

How to slake a planet's thirst

Our growing demand for water threatens the world's development and security. Solving this crisis need not involve flashy technologies. But it will require science, plus a large dose of political will.

By this time next year, some 1.75 million people will have died before their time for the simple reason that they must scratch out their existence without access to safe drinking water. This is the toll from cholera, dysentery and other diarrhoeal diseases that the World Health Organization attributes to unsafe drinking supplies.

United Nations (UN) officials had hoped that this week would hoist the injustice of these avoidable deaths to the top of the world's news agenda, through coverage of the 3rd World Water Forum. Now under way in Japan, the forum is billed as the most important meeting on freshwater resources ever held. But its organizers' plans for a high media profile have been undone by the crisis over Iraq — all in all, this must be a depressing time to work for the UN.

Whatever transpires in the Iraqi desert over the next few weeks, the water crisis will still be with us, quietly exacting its devastating toll (see page 251). As the world's population continues to expand, and climate change alters the hydrology of river basins, as many as 7 billion people in 60 countries could face water scarcity by 2050, if action isn't taken.

So what needs to be done? In large part, the answer lies in the development and large-scale application of relatively simple technologies. Sophisticated solutions such as solar-powered desalination plants have a role to play, particularly in wealthy yet arid countries, but the two biggest components of the global water crisis are the contamination of drinking supplies with human faeces, and the massive wastage of water that is inherent in prevailing agricultural practices. Both of these problems could be alleviated by refining and implementing existing technologies.

Drain on resources

The world's difficulties with sanitation might be eased if we abandon the idea of disposing of our waste with a flush. In the rich North, flush toilets are the largest single drain on domestic water supplies. But there is no need for the developing countries that currently lack adequate sanitation to imitate this wasteful practice, which would require substantial investment in sewage-treatment infrastructure. Ventilated, composting toilets can now safely turn human waste into an odourless, soil-like residue. Further development may be needed to reduce the cost of the technology. But as demands for water grow, this has to be a better solution than a system that involves the contamination and subsequent treatment of precious water supplies.

Meanwhile, irrigation for agriculture accounts for more than two-thirds of humanity's use of water. In many cases, the techniques used are much the same as when the farmers of Mesopotamia first diverted water from the Euphrates some 6,000 years ago: dig a channel from a river to your crops, and let gravity do the work. Thanks to seepage and evaporation, however, this can result in almost 60% of the diverted water being lost. Today, we do not need to be so wasteful. By using 'micro-irrigation', in which water is piped and fed onto crops through sealed systems, irrigation can be made 90% efficient. Much greater use could also be made of recycled waste water, rather than drawing from supplies that could be used for human consumption.

Getting bright young scientists and engineers interested in the world's water problems is vital. One difficulty, in that regard, is that composting toilets and micro-irrigation aren't seen as sexy topics. Yet there are prominent role models to show that scientific excellence and the application of appropriate technologies can go hand in hand. Step forward Rita Colwell, director of the US National Science Foundation, who for the past three decades has studied the factors that influence outbreaks of cholera. Today, her team's research incorporates the latest techniques of genomics and remote sensing, yet Colwell remains aware of the value of low-tech solutions. One of her most recent papers evaluates the use in Bangladesh of folded sari material to filter *Vibrio cholerae*, the cholera pathogen, from drinking water (R. R. Colwell *et al. Proc. Natl Acad. Sci. USA* **100**, 1051–1055; 2003).

Political solution

But it is at the political level that the battle to avert the global water crisis will ultimately be won or lost. For too long, politicians have tried to tackle problems with water resources through centralized control and grandiose engineering projects. 'If it flows, dam it', has been their mantra. These schemes have brought benefits through an expansion of hydroelectric power, and have allowed vast areas to be opened up for cultivation, but the human and environmental consequences have often been disastrous.

Take the Aral Sea, a giant salt lake between Kazakhstan and Uzbekistan in central Asia that is rapidly being turned into a dust bowl, destroying a unique ecosystem and with it the livelihoods and health of many of the region's 5 million people. This environmental catastrophe, today a source of tension between five nations, is a consequence of the former Soviet Union's policy of diverting water from the rivers that feed the Aral Sea to irrigate its vast plantations of cotton.

Such aberrations are not the exclusive preserve of socialist planning — witness Saudi Arabia's draining of its aquifers in the 1980s and 1990s in order to become an exporter of wheat, and the continuing obscenity of sprinklers watering golf courses in the parched western United States. We need to move away from such misguided adventures, which will require more decisions about water resources being taken in close consultation with local water users.

Most fundamentally, however, averting the global water crisis will require developed nations to summon the political will — and hard cash — to help the world's poorest countries tackle water-borne disease, and put their use of water on the path to sustainability. Officially, the world has already signed up to the goal of halving the number of people lacking access to safe drinking water by 2015. But this will cost tens of billions of dollars, and even before the crisis in Iraq consumed their attention, world leaders were showing little commitment to providing the necessary funds.

The 3rd World Water Forum is in danger of becoming "the biggest and most expensive non-event in history", one official with a UN agency confided to *Nature* last week. For the sake of the billion-plus people for whom having a drink of water involves a game of microbiological Russian roulette, things have to change. ■





Growth industry
Alliance promises
Africa affordable
high-tech agriculture
p246



Lucky charm
Quark hunt gives
accelerator new
lease of life
p248



Piece by piece
NASA struggles to
find cause of shuttle
explosion
p249



Steamed up
James Watt
collection goes
under the hammer
p250

Biodefence plans earn lukewarm response from US academics

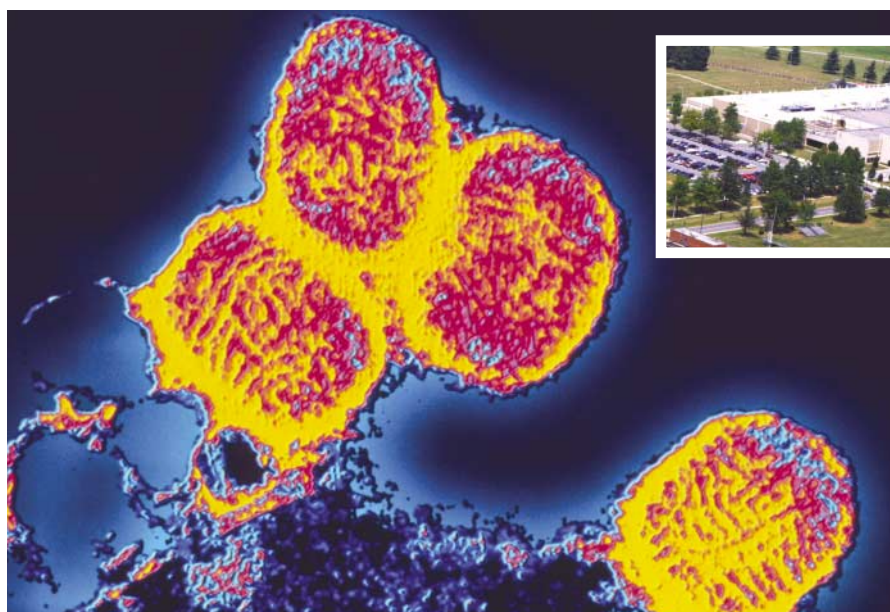
Erika Check, Washington

For a year and a half, biomedical scientists in the United States have heard a steady refrain: get set for the era of biodefence.

Now the federal funds are flowing, and policy-makers are calling for researchers to join the fight against bioterror. But last week the first-ever Biodefence Research Meeting — held in Baltimore, Maryland, by the American Society for Microbiology — highlighted the challenges facing research administrators as they attempt to encourage university researchers to join the gold rush.

Most of the research presented at the meeting came from government facilities such as the Lawrence Livermore National Laboratory in California and the US Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland. These labs have a huge head start in biodefence research, and already have the facilities and culture that promote such work. And their staff researchers dominated the meeting's poster sessions, where more than 200 new ideas lined up for attention.

But observers say that these labs cannot do the whole job by themselves, and that universities must be involved. "We need to try to get the best and the brightest from the broadest spectrum of the scientific community engaged in this effort," says Gail Cassell,



Variola virus is one of the agents targeted by labs such as the US Army Medical Research Institute (inset).

a research scholar in infectious diseases and a vice-president at drug company Eli Lilly in Indianapolis, Indiana.

Cassell was one of those at the meeting who identified areas in biodefence that need urgent attention. These range from basic work on pathogenesis and human immunity to

future applications such as universal vaccines and therapeutics. Perhaps most pressing is the need for new vaccines, antibiotics and antiviral drugs, which will be difficult to develop (see 'On the trail of better treatment', below).

Some at the meeting were asking whether the maze of regulations and security issues

On the trail of better treatment

Brand-new classes of antibiotics and antiviral drugs top most people's list of biodefence priorities, but how far away are they?

Optimists say that HIV is an example of how targeted research can yield results. Twenty years ago, the US Food and Drug Administration had approved only one antiviral drug. But the HIV epidemic fuelled spending in virology, and now more than 30 antivirals are on the market.

One such drug — cidofovir — was initially developed to treat infections in patients with AIDS. But a drug screen revealed that it also combats the variola virus, which causes smallpox.

Cidofovir isn't perfect — it has to be given through an intravenous drip for at least two

weeks in AIDS patients, and the full course of treatment can cause kidney problems. These factors might render it less than effective in confronting a bioterrorist attack. But at last week's Biodefence Research Meeting in Baltimore, researchers showed that these problems might not be insurmountable.

Earl Kern, a virologist at the University of Alabama at Birmingham, presented results from a trial using a modified version of cidofovir that can be taken orally. In mice infected with variola or a closely related virus, Kern found that fewer of the treated mice died than those that were not treated. And Mick Hitchcock, a vice-president at drug firm Gilead Sciences in Foster City, California, showed

that in human patients given a short course of cidofovir, only 3% had serious toxic side effects. "We think this is really the most exciting thing coming along in smallpox," Kern says.

But there are still many potential threat agents for which there are no treatments. And it is possible that a bioterrorist could engineer an agent to be resistant to specific drugs. Researchers at the meeting discussed ways of building universal antivirals and antibiotics to address these issues.

"There are technical and scientific challenges we have to overcome," says Gail Cassell, a vice-president at US drug company Eli Lilly, "and it may be more difficult than any of us appreciate."

Erika Check

A. PASIEKA/SPL; T. JACOBSEN/AP (INSET)

raised by biodefence work will deter academic researchers from joining the field. "The security issues are going to be a new thing for a lot of the academic scientists who want to get involved in this," says Gigi Kwik, a fellow with the Johns Hopkins Center for Civilian Biodefense Strategies in Baltimore, Maryland.

Some of these issues involve regulations on who can work on dangerous pathogens, and where such research can be done. Other researchers expressed fears about the harsh scrutiny that biodefence is likely to attract. Earlier this year, for example, Thomas Butler, a biologist at Texas Tech University in Lubbock was dragged off to prison and publicly humiliated after wrongly accounting for the disposal of some samples of plague bacteria (see *Nature Med.* **9**, 247; 2003).

Another factor causing reticence among academics is apprehension that the government will step in and regulate the kind of data that can be published. Some researchers complain that they were not consulted on a recent publishers' statement on biosecurity and scientific literature signed by this journal, among others (see *Nature* **421**, 771; 2003), and that it may worsen that situation. "There should have been a dialogue, but there wasn't one," says microbiologist Stanley Falkow of Stanford University.

Critics have also claimed that the biodefence programme lacks the emphasis on basic research needed to crack its most important challenges. "This is really a long-term issue that is not going to be addressed in four or five years," says Falkow. Anthony Fauci, head of the National Institute of Allergy and Infectious Diseases, has repeatedly defended his agency's commitment to basic research. But of the \$1.5 billion in new biodefence money that the National Institutes of Health (NIH) will give out this year, \$454 million will go towards work on vaccines, diagnostics and therapeutics, and only \$297 million to basic research.

Despite these concerns, many observers remain confident that the vast sums of money being poured into biodefence will lead to gains against medically important diseases. "There will be a lot that comes from this investment that will have very broad applications, not just to threat agents, but to many other naturally occurring pathogens," says Cassell.

Others suggest that the size of the investment is sure to get academics involved. "As the infrastructure is enhanced, it will become easier and easier for more people to become involved in biodefence," says Gary Nabel, director of the NIH's Vaccine Research Center in Bethesda, Maryland. "I do hope there is interest out there." ■



Future harvest: African farmers are set to get free rights to agricultural tools and technologies.

Biotech firms join charities in drive to help Africa's farms

Hannah Hoag, Washington

An alliance of agricultural biotechnology corporations, charities and aid organizations is pledging to help provide African farmers with the latest technologies at prices they can afford.

The African Agricultural Technology Foundation (AATF) has been set up in Nairobi following an initiative led by the New York-based Rockefeller Foundation. Rockefeller president Gordon Conway says it will help to "unjam the logjam" of intellectual property rights that often prevents poor countries from exploiting advances in agricultural technologies such as plant breeding and transgenics.

Four companies — DuPont of Wilmington, Delaware; Dow AgroSciences of Indianapolis, Indiana; Monsanto of St Louis, Missouri; and Syngenta of Basel, Switzerland — say that they will provide the AATF with patented technologies and technical information without demanding royalties.

Eugene Terry, former director-general of the West Africa Rice Development Association in the Ivory Coast, who will head the new organization, says its priorities will be driven by the demands of Africa's farmers. "We don't want to push any technologies," he says. Farmers' organizations and African governments will be asked to suggest projects for the AATF, which will then help them to negotiate rights to the necessary tools and technologies.

The complex patent arrangements surrounding transgenic crops, as well as the technologies used to produce them, hamper efforts to devise practical solutions to the problems of farmers in poor countries, experts say. But the AATF says it will offer conventional management techniques, as well as new technologies, in response to farmers' problems.



Eugene Terry: pledges to meet farmers' needs.

Conway says he doubts that transgenic crops will be the new foundation's principal contribution, and points to the benefits of less advanced technologies. Researchers, for example, have created banana plants that are free of fungi and parasites by using tissue-culture techniques, and have used genetic markers to develop crops with

increased levels of vitamins and minerals.

"What Africa needs is a diversity of approaches, and the AATF is an important addition to the portfolio," says Calestous Juma, director of science, technology and innovation at Harvard University's Center for International Development. Reliance on traditional farming methods alone will lead the continent to a "fateful destiny of isolation," he warns.

The Rockefeller Foundation has donated \$1 million to plan the operation, and the US Agency for International Development has put in \$550,000. Both will match these contributions this year, and Terry says he is in negotiations with other donors to raise the rest of the AATF's \$2.5-million annual operating budget.

The foundation will not have to pay royalties to the four biotechnology companies. But William Neibur, Dupont's vice-president of product development, warns that some patent arrangements will involve parties that will not necessarily provide intellectual property for free. "There will be royalty-free agreements, but there may also be ones that incur royalties," he says. ■

♦ www.rockfound.org

Health labs focus on mystery pneumonia

David Cyranoski, Tokyo

The World Health Organization (WHO) is coordinating a massive international investigation into a mysterious pneumonia-like disease that may have originated in the southern Chinese province of Guangdong.

On 17 March, the WHO announced that 11 laboratories in 10 countries would collaborate in an attempt to identify the cause of severe acute respiratory syndrome (SARS). The disease, which begins with a fever, cough and shortness of breath, can develop into a pneumonia-like condition that in some cases results in death.

As of 18 March, global health authorities outside China had reported 182 cases of SARS and four deaths across Hong Kong, Canada, Singapore, Taiwan, Germany, Switzerland and Thailand. The WHO took the rare step on 15 March of branding the outbreak "a worldwide health threat".

It remains unclear whether the first reported cases of SARS, in Vietnam and Hong Kong in late February, are related to more than 300 similar cases of atypical pneumonia that began last November in Guangdong and reportedly peaked in early February.

"The clinical picture is very similar, but absolute linkage with the Guangdong cases cannot be made at this time," says Masato Tashiro, director of the WHO's collaborating centre for influenza at the National Institute of Infectious Diseases (NIID) in Tokyo. Tashiro met with health officials in Beijing earlier this month to discuss the Chinese outbreak. But Klaus Stöhr, director of the WHO's global influenza network, says that the timing and clinical similarities "would be too much of a coincidence for the outbreaks not to be linked".

Whether the cause of SARS is bacterial or viral is still unknown. Laboratories initially



Protection is being stepped up at Hong Kong's hospitals after cases of an unusual respiratory syndrome.

looked for stretches of genetic sequence associated with known agents in samples taken from patients. Tashiro, for instance, has been searching for the RNA of the influenza virus in blood serum from patients in Hanoi, using a method called the reverse-transcriptase polymerase chain reaction.

This method converts any viral RNA present into DNA. It then uses primers — short DNA sequences that match the DNA in certain viruses or viral families — to bind to the DNA strand and start copying it, making enough copies for the strand to be identified.

That will only happen if the particular virus is present, and Tashiro's samples did not find the influenza virus. But the RNA could have degraded and become undetectable in the week it took for the samples to reach Tashiro's lab. It is also possible that a virus could have mutated so that the primers cannot detect it.

Tests on samples taken in Hong Kong have identified one type of influenza virus and a respiratory syncytial virus — but

neither has the virulence to explain the outbreak, investigators say. The same tests seem to rule out any significantly virulent influenza, including the H5N1 avian flu virus that infected a family in Hong Kong earlier this year (see *Nature* 422, 6; 2003).

Having ruled out many initial ideas about specific agents, laboratories collaborating in the WHO effort, including the NIID and the Centers for Disease Control and Prevention in Atlanta, Georgia, are undertaking a broader, more systematic search. They will start by using less-specific primers that can identify larger virus families.

Researchers are also delving into records of the Chinese outbreak. China, which has been strongly criticized internally and externally for its failure to provide more information when the outbreak first occurred, has now invited a WHO team to visit Guangdong. The team was expected to arrive there this week. But some health officials are already asking if the delay in access may have contributed to the spread of the disease. ■

Chicken flu races through Dutch poultry farms

Alison Abbott

The Netherlands is this week counting the cost of the latest outbreak of avian flu. The flu, which is sweeping through the country's chicken flocks, has already infected more than 50 farms and resulted in the culling of over 2 million chickens.

The economic implications of the outbreak, which began on 28 February, are severe — the Netherlands is Europe's largest exporter of chickens. But the cost could have been much greater.

Initial fears that the disease was linked to cases of avian flu in Hong Kong (see *Nature* 422, 6; 2003) were quashed early on, thanks to the rapid detective work of a team at

Erasmus University in Rotterdam. The Hong Kong strain, known as H5N1, sparked widespread alarm after it killed a man and infected other members of his family.

Within 48 hours of receiving samples of the virus that caused the Dutch outbreak, the Rotterdam team had identified it as an H7N7 strain, revealing that its genetic characteristics are similar to those of a strain identified in mallard ducks in 2000.

Nevertheless, the Dutch virus can infect humans, and as of 18 March it had caused conjunctivitis in at least 32 people, some of whom have flu-like symptoms. Anxious researchers are not ruling out the possibility that the chicken flu could trigger a human

flu pandemic. But for this to happen, the virus would have to infect someone infected with a normal human flu virus, so that the two strains could swap genes to produce a more virulent humanized virus.

This is thought to be highly unlikely. But to reduce any chances of such co-infection, Dutch doctors are pursuing aggressive drug treatment of those infected, in an attempt to reduce their viral load.

Avian flu outbreaks are rare, but can kill entire flocks in a few days. Transmission of the virus from wildfowl to domestic chickens may have occurred through free-range hens that spend time outdoors, suggests Albert Osterhaus, who heads the Rotterdam team. ■

Revamped accelerator seeks quark's secrets

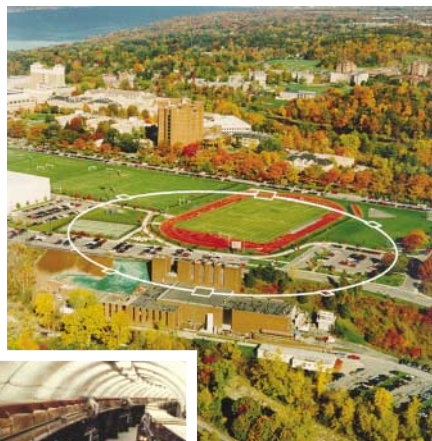
Geoff Brumfiel, Washington

An ageing particle accelerator in New York state is to get a fresh lease of life — hunting down the mysteries of the charm quark, one of the most elusive fundamental building-blocks of matter.

The US National Science Foundation will give Cornell University's particle-physics laboratory \$124 million over five years to pursue the investigation, and to upgrade the synchrotron light source attached to the particle accelerator.

The Cornell Electron Storage Ring (CESR), which sits under a sports field on the university's Ithaca campus, was built in 1979 to examine collisions between electrons and positrons, their antimatter counterparts. For most of the facility's life, high-energy physicists have used these collisions to create another type of quark known as the bottom quark, to study its fundamental properties. "I think it's fair to say that the most information on the bottom quark has come from this facility," says David Cassel, deputy director of Cornell's particle-physics lab.

The planned upgrade to the facility will actually reduce the energy of electrons and positrons in the ring, to around half of its current level of 5 giga-electron volts (GeV). This will be done by devices known as 'wigglers', which act like speed bumps to slow the electrons and positrons down to 2.0–2.5 GeV



Charmed life: Cornell's particle-physics laboratory will now be redirected to studies of the charm quark.

— the right energy, researchers hope, to generate charm quarks.

Charm quarks, which are among the six quarks featured in physicists' standard model, were first observed in 1974 by teams led by Burt Richter, at the Stanford Linear Accelerator Center (SLAC) in California, and Samuel Ting at Brookhaven National Laboratory in New York state. But according to Cassel, "there are a number of crucial things we still don't know about them". For example, researchers have still to obtain precise measurements of the lifetimes of charm-filled particles and the various ways in which they decay.

The revamped facility will also study the strong nuclear force, which binds quarks together to create protons and neutrons, and also keeps those particles together to form atomic nuclei, despite the huge repulsive electromagnetic force between them. In particular, physicists will be searching for 'glueballs', which are clusters of gluons, the particles that carry the strong force. "All theories of the strong force demand that there be glueballs," says Cassel, "but nobody has been able to verify that they exist."

"It's a superb team and the timing is just right," says Persis Drell, a former deputy director of Cornell's particle-physics lab and now associate director for research at SLAC. Drell says that the new investigation will neatly complement Stanford's work on particles known as B-mesons. Many researchers hope that studies of B-mesons will unravel the mystery of why there is more matter than antimatter in the Universe (see *Nature* **419**, 24–27; 2002). "When a B-meson decays, it decays into a charm quark most of the time," Drell explains. "So these measurements will be of value to our experiments."

Some \$25 million of the CESR's new funding will go on upgrading its synchrotron light source, known as the Cornell High Energy Synchrotron Source. The source produces X-rays with which visiting researchers can probe materials and molecules. But because Cornell's particle physicists now plan to run the accelerator at energies that are too low to produce X-rays, they will have to work at different times from the visiting scientists. ■

NATO reform promises more publicity for less science

Quirin Schiermeier, Munich

The North Atlantic Treaty Organisation (NATO) is to cut back its small but highly respected Science Programme and merge it with its public-relations office.



NATO's science division, which was set up in 1958, has supported high-quality military and civilian research. Its goal has been to strengthen scientific links between NATO states, and since the end of the cold war it has concentrated on supporting partnerships with scientists in new member states in central Europe and the 22 NATO 'partner countries' in eastern Europe and mid-Asia.

But the programme's 2003 budget has been cut to US\$20 million, from \$23 million last year. And next year it will be merged with NATO's Office of Information and Press to form a new public diplomacy division.

The move worries some researchers, who fear the programme may be heading for oblivion. "The situation seems stable for the coming year, but I am not very optimistic about the longer-term future of the Science Programme," says Charles

Buys, a geneticist at the University of Groningen who represents the Netherlands on NATO's Science Committee, which was informed of the planned change when it met in Brussels on 6 March.

But Brian Heap, who represents Britain on the committee, says that the new structure could actually raise the profile of NATO's research. "It is crucial to bring to public attention the fact that NATO is promoting peace and solidarity, while tackling the humanitarian challenges in its partner countries," he says.

NATO plans to continue supporting partnerships that involve researchers in the new member states, but they will be reduced in number and limited to research topics directly relevant to NATO's mission, such as counter-terrorism. ■

♦ www.nato.int/science

NASA intensifies probe into Columbia failure

Tony Reichhardt, Washington

"It's a detective story we are taking one piece at a time," says John Barry, a major general in the US air force and member of the board seeking the cause of the Columbia tragedy. But six weeks into the investigation, Barry and other board members admit that the reason why the space shuttle exploded on 1 February remains elusive.

At the same point after the 1986 Challenger disaster, investigators had already zeroed in on the cause. By contrast, those studying the break-up of Columbia are still collecting debris and running basic tests on the parts that could have failed.

Clues gleaned from flight data, recovered debris, and pictures and videos of the break-up have, however, allowed the investigators to construct a detailed timeline of the beginning and end of Columbia's last voyage.

Data from the booster rockets show that 62 seconds after launch, wind turbulence briefly exerted a stronger-than-usual force on the shuttle's left-hand side. Launch videos show that after a further 19 seconds, three pieces of what was probably the tank's foam insulation, but could have been foam mixed with ice, appeared to strike the leading edge of Columbia's left wing.

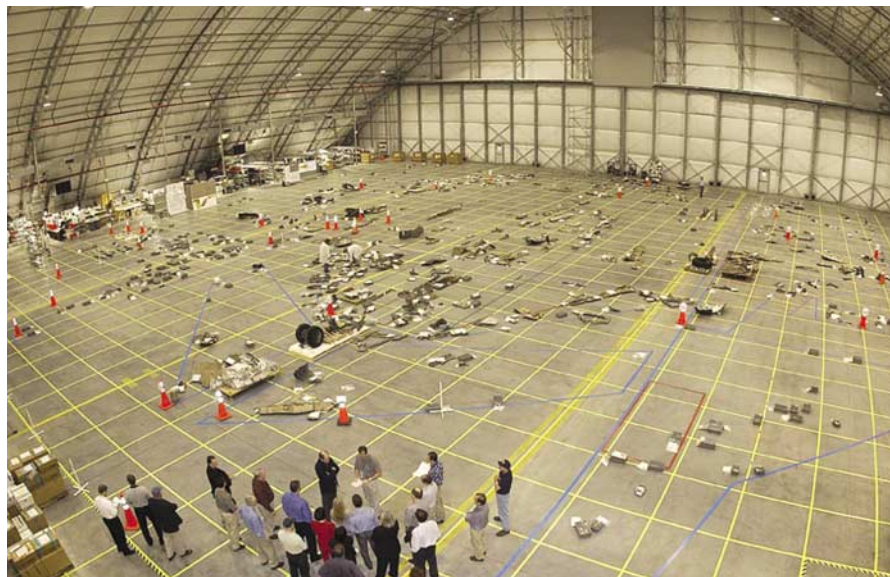
Investigators say that when the shuttle re-entered the atmosphere 15 days later, hot gases entered the left wing. These acted like a blowtorch, melting some of the aluminum frame and causing widespread damage.

But connecting the launch-day impact to those final events has not been straightforward. The debris struck several of 22 desk-sized panels of reinforced carbon insulation around the wing's leading edge. However, no panels are seen falling off Columbia in the launch videos, and the 200 or so pieces of the left wing recovered to date show no sign of having been hit by a large piece of debris. And wind-tunnel tests suggest that losing one of the panels would not be enough to destabilize the shuttle during re-entry.

Less obvious damage to the panels, or to the insulation that fills the gaps between them, may still be the culprit. Accident investigators have asked researchers at the Southwest Research Institute in San Antonio, Texas, to use high-speed guns to fire pieces of foam insulation at sections of insulation and supporting structure to characterize the kind of damage that this might cause.

The Wright-Patterson Air Force Base near Dayton, Ohio, is generating radar signatures for items such as heat-resistant tiles from the underside of the shuttle and pieces of carbon insulation. These will be compared with radar images of a mysterious object that was seen floating near the shuttle during its second day in space.

Staff at radar stations in southwestern



Accident investigators at the Kennedy Space Center piece together debris recovered from the shuttle.

states are also checking whether signals collected as Columbia flew overhead on its way to the landing strip at Kennedy Space Center in Florida might provide clues as to what went wrong. Ground-based images have shown that at least 16 pieces of debris were shed by the struggling vehicle as it sped over California, Arizona and New Mexico.

Board chairman Harold Gehman, a retired US Navy admiral, hopes to move soon from collecting information to a testing phase. Investigators are deciding what chemical and metallurgical tests to conduct on recovered debris to assess what damage they sustained. More than 28,000 pieces have been recovered to date. "Just laying them out on the floor is not enough," says Gehman.

The search for debris has cost \$138 million so far, with some 4,000 people and more than a dozen aircraft in the field every day. Twenty federal agencies are involved in the

search, which will become more difficult in the coming weeks as spring comes to Texas, covering the main search area with foliage.

Meanwhile, NASA staff have begun work on plans to resume flying this year, perhaps as early as autumn. "The best therapy we can have is to be extremely engaged in solving this particular problem," said Ron Dittmore, manager of the space-shuttle programme, at a hearing earlier this month. The plans include reviewing methods for applying insulation to the shuttle's external fuel tank.

NASA administrator Sean O'Keefe said last week that the agency is not prejudging what Gehman's board will find, nor is it assuming that the shuttle will be cleared to fly this autumn. But NASA wants to do what it can now to address concerns that have already come up in the investigation. "The alternative is to sit here on our hands and wait for a report to be released," said O'Keefe. ■

Alien-hunters get scope to search

John Whitfield

The most promising results from the SETI@home search for extraterrestrial life are to be followed up using the world's largest radio telescope.

Since May 1999, SETI@home — run by the Search for Extra Terrestrial Intelligence, a non-profit organization — has divided radio signals from our Galaxy into quarter-megabyte chunks and sent them to the desktop computers of more than four million registered volunteers. The computers flag up strong or regular signals for subsequent analysis.

Later this month, researchers from the University of California will use the Arecibo telescope in Puerto Rico to study the locations indicated by the 150 most promising results. The project's chief scientist, Dan Werthimer of the University of California, Berkeley, says that there is about a 1-in-10,000 chance of finding something.

SETI staff say they could be looking at the wrong wavelengths or in the wrong place. "I'm optimistic that the Universe is teeming with life," says Werthimer. "But I'm not convinced we know exactly how to find it." ■

♦ <http://setiathome.ssl.berkeley.edu>

Danish study hints at biodiversity boost from transgenic crops

Copenhagen Tentative evidence that some genetically modified crops could benefit biodiversity has emerged from a trial of transgenic beet in Denmark.

Researchers from the National Environmental Research Institute in Silkeborg compared levels of weeds, insects and other animal life in fields of conventional beet with those in a strain that was genetically engineered to be resistant to a glyphosate herbicide. After application of the herbicide, more weeds and insects survived in transgenic beet fields, although beet yields were not cut.

But the study does not prove that genetically engineered crops will not harm the environment, warns Marianne Pedersen, one of the Silkeborg researchers. "One-year positive results may be outweighed by longer-term drawbacks," she says, "because weed in glyphosate-treated sites is killed before it can produce seeds in late summer." Such uncertainties may be cleared up by Britain's farm-scale transgenic beet study, the results of which are due out this year.

Crop trials elsewhere are proceeding less smoothly. Plant scientists at the Swiss Federal Institute of Technology in Zurich last week saw a proposed study of transgenic wheat cancelled when the country's highest court ruled that they had not taken environmental objections into consideration.

Public starts to accept use of animal experiments...

London Good news for those advocating the use of animals in medical research — British public opinion may be swinging their way.

UK poll specialists MORI asked whether animal experimentation was always, sometimes or never justified, and scored respondents' answers + 1, 0 or - 1, respectively. In the case of research on diseases such as AIDS, the average score jumped from 0.08 in 1999 to 0.29 in 2002.

The use of animals in studies of other life-threatening diseases and in basic biological research also gained acceptance, and respondents reported more faith in the



Animal pharm? British opinion may be moving towards using animals for medical research.

Watt's lots of steam

London A selection of James Watt's possessions, from an early photocopier to his description of how he was drawn to steam power by a boyhood observation of his aunt's kettle, are due to be auctioned at Sotheby's in London this week.

Watt's greatest achievement was a steam engine that was four times more powerful than the best of the day. But he also invented an early chemical photocopier, an example of which from 1790 is in the sale, along with his desk, library, telescope, dinner service and a model steam engine (right).

The sale is expected to raise more than £500,000 (US\$800,000). The auction follows the death last year of Lord James Gibson-Watt, the inventor's great-great-grandson.



regulatory system. The survey was commissioned by the Coalition for Medical Progress, an umbrella group of funding agencies, charities and companies involved in medical research.

Some work on communicating with the public still needs to be done. When asked to name an animal most commonly used in experiments, 2% of respondents suggested the horse, whereas 12% argued that the use of bacteria in medical research is unacceptable.

...but activists force hotel to ditch meeting

London Protests from animal-rights activists have forced the cancellation of a major conference for technicians working with laboratory animals.

The Citywest Hotel near Dublin in Ireland cancelled the booking for the Institute of Animal Technology's annual congress, due to be held there next month, after pressure from the campaigners. The institute says that the hotel's decision followed "attacks on its facilities, staff and guests by anonymous animal-rights extremists", although the hotel would not comment.

Some members of the institute are staff at Huntingdon Life Sciences, an animal-testing facility near Cambridge that was targeted for closure by a protest group called Stop Huntingdon Animal Cruelty. The group's website tells of the campaign against the hotel and now claims victory.

A spokesperson for the institute says that the congress will take place later this year at a different venue.

Plant breeders see benefit of growth in genomics

Munich The ability of the genomics revolution to boost conventional plant breeding is to be explored by science academies from European Union member states.

Transgenic technologies have captured the

bulk of crop-research funds in recent years at a time when many conventional plant-breeding programmes are suffering from cutbacks (see *Nature* 421, 568–570; 2003). But advocates of conventional approaches point out that genetic modification involves the addition of individual genes, whereas conventional breeding can change several traits simultaneously.

Francesco Salamini, a plant scientist at the Max Planck Institute for Plant Breeding Research in Cologne, Germany, says that the use of genomic data could speed up conventional approaches. Researchers must currently grow plants to see if they have acquired a particular trait, but sequence data could reveal the presence of identifiable and easy-to-check genetic markers that are linked to superior versions of useful traits.

Clonaid wins 'snake oil' for clone claims

San Francisco With tongues firmly in cheeks, experts on ageing last week named Clonaid, the company that claims to have cloned humans, as a winner in the second annual Silver Fleece Awards. The prize, a bottle of vegetable oil labelled "snake oil", recognizes the year's most outrageous claims about human ageing.

Clonaid, which was founded in 1997 by a religious sect that believes that aliens populated the Earth through cloning, claims that "cloning will enable mankind to reach eternal life". Many scientists argue that the company's publicity has accelerated a move by Congress to ban all forms of cloning in the United States, despite its medical potential.

Clonaid officials accuse the award's originators, which include Leonard Hayflick, a biologist at the University of California, San Francisco, of lacking vision. A second award went to Urban Nutrition, which markets an anti-ageing potion, called Longevity, on the Internet.

The world's forgotten crisis

Over a billion people cannot get clean water, and things are getting worse.

This week in Japan, some 10,000 delegates from across the globe will meet to examine what, for many of the world's poorest people, are issues of life and death. At sessions in Kyoto, Shiga and Osaka, the 3rd World Water Forum will highlight a crisis that currently forces more than one billion people to drink from sources contaminated with human waste, and leaves countless millions more with insufficient supplies to water their crops, or to spur industrial development. The meeting is the culmination of the International Year of Freshwater 2003, declared by the United Nations to raise awareness of the worsening state of the world's water resources.

Given that goal, the forum could hardly have been held at a less opportune moment. Events in Iraq may limit media coverage of the 3rd World Water Forum to mere footnote status. "Our discussions will have far more effect on humankind in the twenty-first century than the current crisis in the Middle East," claims William Cosgrove, vice-president of the World Water Council, a think-tank dedicated to improving water resources. But the forum's organizers must know that there is little hope of gaining firm commitments from nations that are focused on war.

Cosgrove may nevertheless be right. The water crisis is real. If action isn't taken, millions of people will be condemned to a premature death. According to the *World Water Development Report*, a UN survey

prepared for the forum, population growth, pollution and climate change are conspiring to exacerbate the situation. Over the next two decades, the average supply of water per person will drop by a third. Heightened hunger and disease will follow. Humanity's demands for water also threaten natural ecosystems, and may bring nations into conflicts that — although they may not lead to war — will test diplomats' skills to the limit.

In the following pages, *Nature* summarizes the global water crisis through a series of maps and graphics, and then uses the case study of Bangladesh to reveal the scientific and political complexities involved in addressing a nation's requirement to be supplied with — and sometimes protected from — water.

Numbers alone can never tell the whole story, but our graphical summary helps to emphasize the intimate link between poverty and the water crisis — it's no coincidence that water-borne diseases impose a far greater burden on sub-Saharan Africa than anywhere else. And as our case study emphasizes, science must have a frontline role in addressing the issues involved. Without

the skills of specialists such as hydrologists, meteorologists and epidemiologists, there can be no workable solutions.

But although necessary, scientific expertise is far from sufficient. "Inertia at the leadership level, and a world population not fully aware of the scale of the problem, means we fail to take the needed timely corrective actions," the *World Water Development Report* concludes.

Until now, the report charges, there have been too many empty promises. In 2000, as part of the UN Millennium Development Goals, governments agreed to halve the proportion of people without access to safe drinking water by 2015. The pledge was reinforced at last year's World Summit on Sustainable Development in Johannesburg, South Africa. But on current form, there is no chance of it being met.

Aware of the danger of becoming yet another talking shop, the organizers of the 3rd World Water Forum have asked delegates to come ready to make concrete commitments to measures designed to help solve the water crisis. The question is: when the smoke from the current confrontation in Iraq finally clears, will the world remember to hold the forum's participants to account? ■

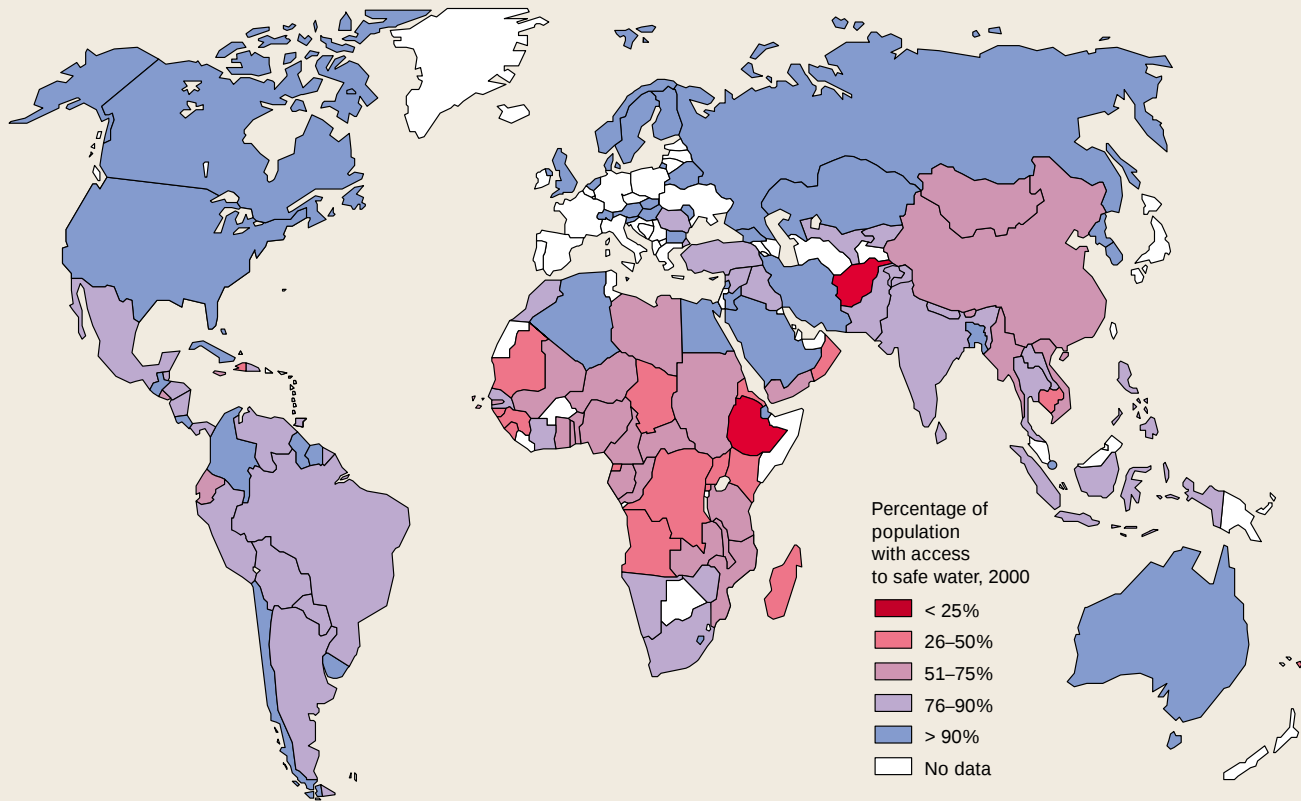
Peter Aldhous, chief news and features editor.
International Year of Freshwater 2003

♦ www.wateryear2003.org

3rd World Water Forum

♦ www.world.water-forum3.com



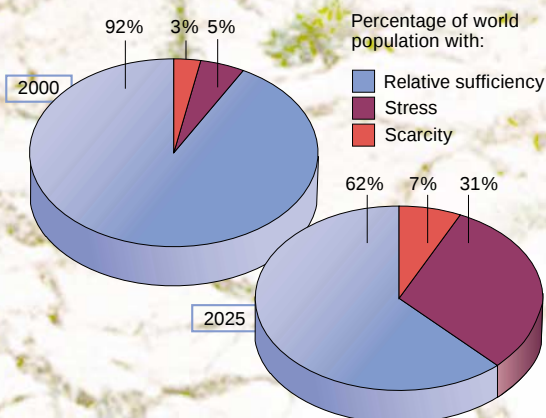


Atlas of a thirsty planet

Red is for danger, and in these maps it reveals where the water crisis is biting hardest. Above left, we show country-by-country figures for the percentage of people with access to safe water. Above right, we reveal the percentage of the total disease burden for each country that is caused by unsafe water, poor sanitation and hygiene. It's not just drinking supplies that we need to be worried about: the pie charts below show that water — clean or contaminated — will become an increasingly scarce commodity. Stress is defined as less than 1,700 m³ per person, per year. Scarcity means less than 1,000 m³ per person, per year. Research and text by *Nature* intern Hannah Hoag. We are grateful to Peter Gleick of the Pacific Institute for Studies in Development, Environment and Security, Oakland, California, for his assistance in collating these data. See www.nature.com/nature/focus/water for enhanced and extended versions of these graphics.

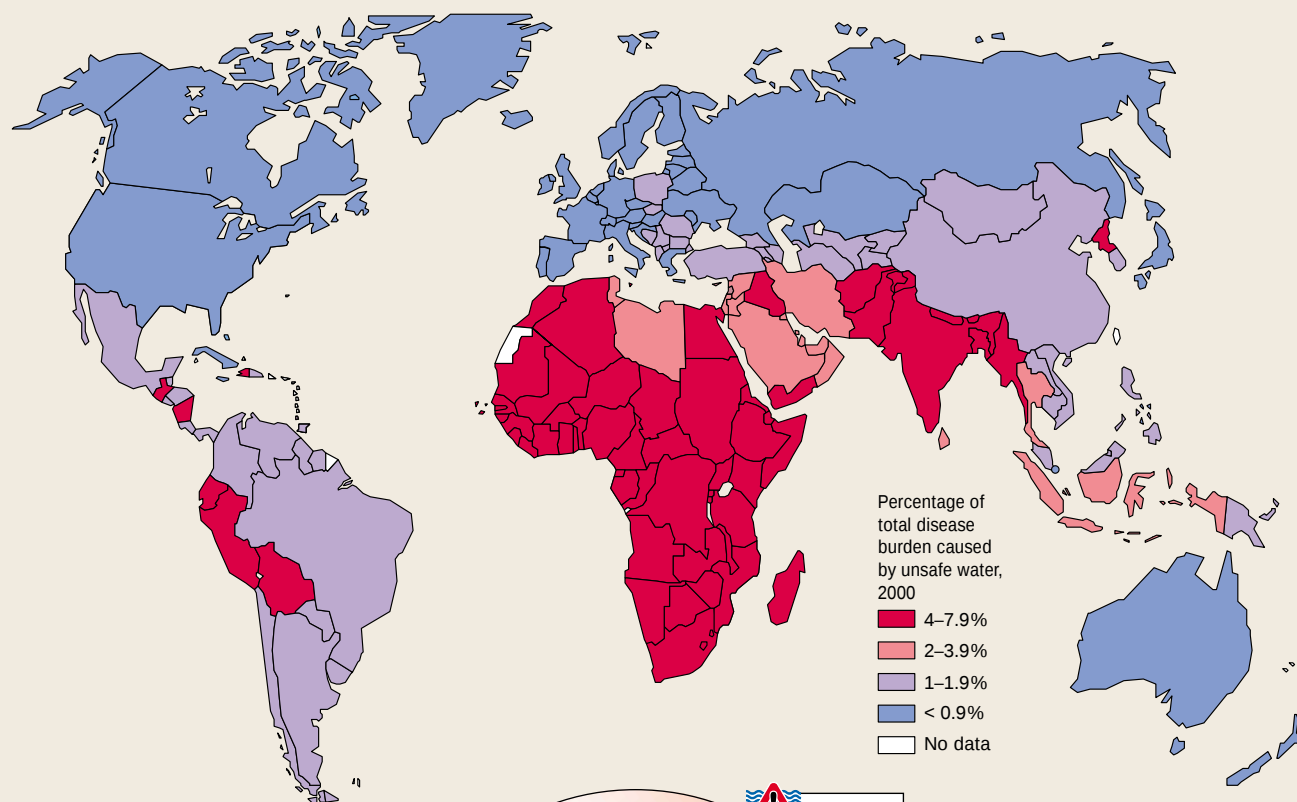
The snow-covered peaks of the Sierra Nevada and Rocky Mountains act as water towers for the American west. Each annual snowmelt recharges the region's river basins and reservoirs. But climate change could bring hydrological chaos. Average temperatures in the region will rise at least 2 °C over the coming century, bringing more rain, less snow and earlier melting. This would halve the snowpack's volume, increasing flood and landslide hazards in the winter and spring. In the autumn, flows would reduce, allowing salt water to penetrate downstream estuaries.

Running dry



Belgium, according to the *UN World Water Development Report*, is bottom of the world rankings for water quality. The rankings do not simply reflect the safety of water supplies — you're definitely better off drinking tap water in Brussels than in Burundi — but rather are a weighted composite of ratings of various aspects of water-resources management. Heavy industrial pollution and agricultural run-off, compounded with inadequate wastewater treatment and balkanized government responsibility for the issues, explain Belgium's dismal showing.

...AND EFFECT



Flashpoints for conflict



The North China Plain is drying up, and inefficient agriculture is to blame. The region, drained by the Yellow, Hai and Huai rivers, supports more than half of China's wheat crops, a third of its maize production, and extensive plantations of cotton and tobacco. Farmers are pillaging the rivers and digging ever deeper into aquifers. In 1997, the Yellow River ran dry for 226 days; between 1991 and 1996, water tables beneath the basin dropped by an average of 1.5 metres, and now China's food security is threatened: since 1998, grain production has been dropping steadily.

Local difficulties

Civilizations have relied on the Tigris–Euphrates basin for more than 7,000 years. Today, Turkey, Iraq and Syria depend on the rivers' water and hydroelectric potential. And tension is rising between the first two countries. Turkey's Southeastern Anatolian Project — an ongoing plan to install 22 dams and 19 hydroelectric plants along the Euphrates and the Tigris — has already nearly doubled the cotton output of the region. But that has come at the expense of Iraq's access to the rivers' life-giving flows.

Delta blues

Living on the flood plain of three great rivers, the people of Bangladesh endure floods, drought, water-borne disease and much else besides. Can they entertain any hope of relief? Tom Clarke investigates.

If Bangladesh were to count her blessings, they would number three: the Brahmaputra, the Meghna and the mighty Ganges. These great rivers are practically Bangladesh's only natural resources. In a predominately rural country in which agriculture and freshwater fishing are the linchpins of the economy, the rivers are the people's lifeblood.

But these blessings, allied with the region's summer monsoon climate, are also a curse. Although almost two metres of rain fall on Bangladesh each year, more than two-thirds arrive in just four months. For much of the year, the vast delta formed by the three rivers is parched, but in many summers their banks burst, causing massive floods. Lacking proper sanitation and water-storage facilities, Bangladesh is also prone to epidemics of water-borne disease. "Even during floods, the major problem is the availability of safe water," says environmental scientist Atiq Rahman, executive director of the Bangladesh Centre for Advanced Studies in Dhaka.

Climate change will only make matters worse, with shifting patterns of rainfall and rising sea levels threatening to render large tracts of agricultural land useless. Add a cruel and recently discovered twist — the poisoning of many millions of people by well water contaminated with arsenic — and it's clear that Bangladesh represents a challenging case study for anyone who wants to solve the world's water woes.

"Nowhere is water more dominant in people's lives," says John Soussan, a geographer at a branch of the Stockholm Environmental Institute at the University of York, UK, who studies water issues in Bangladesh. And nowhere better illustrates the complexity of producing workable solutions to water-resource problems. The threats that face



Water, water everywhere: but managing the Bengal Delta (pictured here from space) is far from simple.

Bangladesh's supplies are intimately interconnected: thwart one, and you can create problems downstream — sometimes literally.

Political pressures

Worse still, the issues are overlaid by fractious regional politics. For decades, Bangladesh has been in dispute with its neighbours — particularly India — over their management of the rivers that drain into Bangladeshi territory. For the country's politicians, water is the defining issue, and the Ganges — known in Bangladesh as the Padma — whose basin is home to some 400 million people, is a perennial bone of contention. Despite some progress in reaching agreement over the river's management, Bangladesh still blames India for holding back too much water in the dry season, and letting the floodgates open each time the monsoon threatens.

The picture looks bleak, but experts point out that Bangladesh is, in some ways, a victim of its own success. Given the hydrological hand they were dealt, the inhabitants of the Bengal Delta traditionally grew low-yielding but flood-tolerant rice, and fished wetlands and pools that were recharged by annual floods. This could support a modest population at subsistence levels, but no more. Since the late 1950s, however, aid-donor-backed irrigation schemes, later incorporating

groundwater pumping, have opened up vast areas of fertile delta soil to the plough. Now almost all of the land in Bangladesh that is suitable for agriculture is used. High-yielding rice varieties have boosted productivity hugely, while the development of coastal areas for shrimp farming has also provided further food and revenue.

As a result, the population has quadrupled since 1950. Today, an average of 920 people crowd into each square kilometre of Bangladesh, making it one of the most densely populated countries in the world. Therein lies the problem: population pressure has helped to make droughts more severe and floods potentially more devastating. And with sanitation still inadequate, the rapid population growth of the past half-century has exacerbated problems with water-borne disease. Although the analogy won't endear him to many of his compatriots, Rahman likens Bangladesh to a giant toilet that is adequately flushed just once a year.

By the 1970s, it was clear that something had to be done, and aid agencies — led by the World Bank and UNICEF — hit on the idea of sinking tube wells into the aquifers that lie beneath the delta's surface. Local people, fearful of this subterranean source, initially called it 'the Devil's water'. But when rates of diarrhoeal disease halved, the programme was deemed an unqualified success. By the

1990s, as many as 10 million wells had been sunk, many of them by local companies.

If only things were that simple: the tragedy that has subsequently unfolded reveals in stark terms how 'solutions' to water-resource problems can go spectacularly awry if our knowledge of a region's hydrology and geology is incomplete. It was Dipankar Chakraborti, an epidemiologist in the neighbouring Indian state of West Bengal, who first raised the alarm. In 1988, on a visit to his Bengali parents' rural village, Chakraborti noticed that many local people were suffering from skin lesions and cancers that seemed to be consistent with arsenic poisoning. When he tested samples of well water in his laboratory at Jadavpur University in Calcutta, it became clear why: the villagers' water supply was massively tainted with the metal.

Drinking problem

Over the ensuing five years, Chakraborti and his colleagues at Dhaka Community Hospital showed that the problem extends across large areas of the Bengal Delta on both sides of the India–Bangladesh border¹. Some wells contain 400 times the World Health Organization (WHO) safe drinking-water standard for arsenic. Current estimates are that 80 million Bangladeshis are at risk, with 30 million drinking water containing five times the WHO arsenic limit. "The danger is very, very real," says Chakraborti.

The arsenic was washed from the sediments of the Himalayas, and is thought to have been accumulating in the silt beneath the Bengal Delta for at least 2 million years². The puzzle is why it is now being drawn to the surface in some wells, but not in others. One leading theory is that the arsenic is released from the sediments into groundwater under oxygen-free, reducing conditions. And some researchers suspect that rotting vegetation in the uppermost 30 metres or so of sediment creates just such an environment. That would help to explain why the problem seems worse in shallower wells, and adds to hopes that it may be possible to identify and selectively shut down those that are hazardous.

But recent research from a team led by Shafiqul Islam, a Bangladeshi hydrologist now at the University of Cincinnati in Ohio, suggests that the pumping of groundwater for irrigation seems to be drawing arsenic into deeper wells³. The mechanism remains unclear, but Islam suspects that deep groundwater is being replaced by surface water that is rich in organic material, which then mobilizes previously insoluble arsenic. Chakraborti has also found that wells that his group tested and marked as safe had become dangerous when surveyed again a few years later.

What's more, drinking water may not be the only hazard: rice crops irrigated with arsenic-contaminated groundwater have now been shown to accumulate the toxic element⁴.



Although it's too early to tell whether this poses a serious threat, the finding has sown seeds of doubt about the use of aquifer water for irrigation, which is the mainstay of Bangladesh's agriculture.

Research and new technology will be central to tackling the arsenic problem. Surveying and screening wells is a top priority, yet arsenic-testing kits for use in the field are still unreliable⁵. More work is also needed to confirm the conditions under which arsenic is mobilized from the sediments, and at what depths this occurs.

Glut and dried: Bangladesh's monsoon climate has locked the country into a cycle of flooding (below) and drought (inset), which has been exacerbated by its own water-management efforts and those of its neighbour India.



Even if the arsenic issue can be resolved, that still leaves the difficulty of living with Bangladesh's annual cycle of flood and drought. The floods can be lethal on a massive scale: those of 1974, together with the famine and water-borne disease that followed, killed 30,000. Such tolls led to the construction of nearly 7,000 kilometres of embankments to contain the delta's rivers. But they couldn't prevent 55% of the country from being inundated in 1988, when 2,000 people died.

What has followed represents a small victory for science-based watershed management. Initially, the Bangladeshi government stuck to a plan that relied on further huge engineering projects. But by the late 1980s, academic scientists at the Bangladesh Centre for Advanced Studies and elsewhere were revealing the pitfalls of relying too heavily on engineering solutions. Although embankments may stop floods, they also prevent water from dissipating when they are breached, delaying recovery from floods and ruining harvests that might otherwise have survived temporary submersion. The embankments were also preventing waters laden with fish fry from replenishing the country's wetlands⁶. "In some cases the infrastructure caused more problems than it solved," says Matthew Chadwick, a colleague of Soussan at the University of York.

After "a hell of a lot of convincing", the government was persuaded to rethink its strategy, says Rahman. While still relying on

engineering solutions to protect major population centres, the focus elsewhere was shifted to evacuating people from areas that were threatened with temporary inundation.

This revised approach seems to be working better. When the worst floods in Bangladesh's history struck in 1998, submerging two-thirds of the country, 22 million people were displaced, but just 700 lost their lives. "People are learning to live with floods," asserts Saleem Huq, formerly an adviser to the Bangladeshi government on the environment and now based in London at the International Institute for Environment and Development.

Dam nation

This may be so, but less progress has been made in tackling the other side of Bangladesh's cycle of hydrological poverty and excess. The flow of the country's rivers naturally declines during the height of the dry season, from March to May. But for the Ganges, the problem has been exacerbated by India's construction of the Farakka Barrage, 20 kilometres upstream of the Bangladeshi border. Completed in 1974, the dam diverts water from the Ganges into the Hoogly River, which feeds Calcutta. Following its construction, dry-season flows of the river through Bangladesh were drastically curtailed (see table, below).

The first trials of the barrage poisoned Indian-Bangladeshi relations. It took three years, and several appeals by Bangladesh to the United Nations, to negotiate a five-year water-sharing deal. Such temporary agreements persisted until 1988, when talks broke down amid renewed hostility. Hope was revived in 1996, when the two countries signed a 30-year Ganges Water Treaty, which requires both countries to monitor flows and negotiate shares in the river's water the during the dry season.

Yet despite the treaty, it is thought that less water flows through southwestern Bangladesh in the dry season now than it did in the 1970s. As a result, the Ganges dumps more of its silt in the channels of the delta, rather than in the Bay of Bengal. "It's like cholesterol getting into your arteries," observes Islam. This symptom has further reduced flow in some regions, creating a vicious cycle of hydrological deterioration.

The curtailed flow is also causing problems in Bangladesh's coastal zone, which forms a porous buffer between the Bay of Bengal and the fertile delta soils. As the flow of fresh water falters, salt water is advancing to replace it in the southwest corner of the country, downstream of the Farakka Barrage. In conservation terms, it couldn't



Water marks: skin lesions, a symptom of arsenic poisoning, are rife among the delta's inhabitants.

be happening in a worse place. That region hosts the Sunderbans — the largest expanse of mangrove forest in the world — which provide habitat for the endangered Bengal tiger (*Panthera tigris tigris*), estuarine crocodile (*Crocodylus porosus*) and Indian python (*Python molurus molurus*), as well as 260 species of birds. Already, the tops of the forest's eponymous Sundari trees (*Heritiera fomes*) are beginning to die back — a telltale sign of excess salinity.

Agriculture and drinking water are also threatened. Around the town of Khulna on the fringes of the Sunderbans, well water is now too salty during the dry season for irrigation and drinking. The advancing salt water brings with it marine plankton, including miniature crustaceans known as copepods. Although harmless in themselves, the copepods' body surfaces are the preferred home for *Vibrio cholerae*, the bacterium that causes cholera — and saltwater intrusion has already been linked to unusually severe outbreaks of the disease in Bangladesh⁷.

Climate change could make matters worse still. Recent research by climatologist Peter Webster at the Georgia Institute of Technology in Atlanta predicts that an El Niño-like phenomenon that causes fluctuations in sea level in the Bay of Bengal will intensify. This, coupled with even a slight rise in global sea level, could seriously accelerate saltwater intrusion. At the same time, the cyclones that batter the Bangladeshi coast may become more intense, as will the monsoon rains⁸ — leading most experts to suspect that Bangladesh's flood-drought dichotomy will become more pronounced.

Faced with its current difficulties and the prospect of worse to come, experts agree that

the solution lies in 'integrated water-resource management' — a comprehensive plan that takes account of all of Bangladesh's interconnected water problems. "There needs to be clarity of concept, not just a bunch of random projects," argues Rahman.

Emerging scientific tools should help. Last month, for instance, at the annual meeting of the American Association for the Advancement of Science in Denver, Colorado, Webster unveiled a monsoon model that provided accurate 25-day forecasts of rainfall over the entire Ganges basin in 2002. This model should provide more time for emergency planners to house displaced people during the most severe floods. "Farmers can make informed decisions, too, either harvesting early or planting late," says Webster.

There is even a draft National Water Management Plan on the table. Commissioned by the Bangladeshi government from the infrastructure consultancy firm Halcrow and completed in 2001, the plan calls for an integrated approach to water management. It also recommends greater decentralization to hand more decisions on resource allocation over to local governments who are within earshot of the needs of water users. It's a radical plan, and not without its critics, but many experts agree that it would be a significant step in the right direction.

The stumbling block, as is so often the case, may be at the political level. For a start, the plan was commissioned by a previous administration, and the new government has yet to embrace it. More fundamentally, the goal of decentralization may not yet be practical. "There are huge gaps in the system of local government," says Soussan. In short, Bangladesh needs help in training a generation of technocrats to take ownership of its water-resource problems.

For experts who have devoted their professional energies to studying Bangladesh's water problems, it would be an unforgivable tragedy if these obstacles proved to be insurmountable. "There's enough knowledge out there to address the problems," concludes Chadwick. "Now we really need to start putting some of these ideas into practice." ■

Tom Clarke works in Nature's news syndication team.

1. Chowdry, U. K. et al. *Environ. Health Perspect.* **108**, 393–397 (2000).
2. Kinniburgh, D. G. & Smedley, P. L. (eds) *Arsenic Contamination of Groundwater in Bangladesh*. Br. Geol. Surv. Tech. Rep. WC/00/19 Vol. 2 (2001).
3. Harvey, C. F. et al. *Science* **298**, 1602–1606 (2002).
4. Meharg, A. A. & Rahman, M. M. *Environ. Sci. Technol.* **37**, 229–234 (2003).
5. Rahman, M. M. et al. *Environ. Sci. Technol.* **36**, 5385–5394 (2002).
6. Datta, A. (ed.) *Planning and Management of Water Resources: Lessons From Two Decades of Early Implementation Projects, Bangladesh* (Univ. Press, Dhaka, 1999).
7. Pascual, M., Rodó, X., Ellner, S. P., Colwell, R. & Bouma, M. J. *Science* **289**, 1766–1769 (2000).
8. Palmer, T. N. & Räisänen, J. *Nature* **415**, 512–514 (2002).

Draft National Water Management Plan for Bangladesh
 ♦ www.warpo.org/dnWMP/dnwmp.htm

Mean monthly flows in the Ganges (m³)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1934–74	3,090	2,668	2,287	2,031	2,176	4,489	17,290	38,348	36,063	17,870	7,091	4,180
1974–88	1,932	1,482	1,155	1,063	1,450	3,569	20,111	40,183	39,233	16,685	5,730	2,493

Venezuela: the other side of the story

No political campaigning in public institutions, but free speech is manifest in e-mail debate.

Sir — As members of the board of directors of the Venezuelan Institute for Scientific Research (IVIC), we feel obliged to clarify the following points concerning the Correspondence letter “Venezuelan researchers call for international help” (*Nature* **421**, 473; 2003) by members of the Association of Investigators of IVIC (AsoInIVIC).

First, AsoInIVIC is a private scientific association of about 130 IVIC employees. The total number of IVIC researchers is 124, including 19 postdoctoral fellows. There are also 226 professional research associates, 201 postgraduate students, 242 general staff and 310 administrative support staff. Thus, AsoInIVIC members constitute only a small fraction of the approximately 1,100 people working and studying at IVIC. The sentiments expressed in their Correspondence do not represent the official opinion of IVIC, a public institution; rather, they are the exclusive opinion of AsoInIVIC, which is a private association.

Second, freedom of speech in IVIC is manifest, among other activities, by the

intense exchange of internal e-mails, covering a vast range of subjects from personal to institutional. The laws and regulations concerning the expression of political ideas in Venezuela's public institutions allow the free expression of political opinions, but restrict political campaigns (proselytism). The distribution of the AsoInIVIC letter within IVIC is a regrettable example of unrestrained political proselytism. Equally regrettable is the lack of scientific spirit manifest in its many inaccuracies concerning both IVIC and some other public institutions, for example the Supreme Court of Justice and the ministries of Science and Technology, Environment, and Energy and Mines.

Third, December and January were months of high political activity in Venezuela. During this period a partial strike, difficult to distinguish from a partial lock-out, took place. As can be expected, this made it difficult to accomplish the scientific research that is usually carried out at IVIC. However, during this time most employees regularly

attended and carried out their work.

Finally, it is widely known that the maintenance and development of scientific laboratories of a high international standard represents a considerable cost to our nation. This cost is particularly heavy in the current period of economic difficulty. Under such circumstances, we consider even more deplorable the misuse of our facilities and work time for political propaganda activities.

The board of directors of IVIC comprises seven members, five of whom are signatories of this letter. The other two members (the director and deputy director of IVIC) do not agree with the content, and consequently have not signed.

M. García Sucre Representative of the Ministry of Science and Technology

L. Marciano Representative of the Ministry of Education, Culture and Sports

R. Padrón Representative of the Ministry of Science and Technology

J. Acosta Labour director

L. Burguillos Labour director
Venezuelan Institute for Scientific Research,
PO Box 21827, Caracas 1020A, Venezuela

Venezuela: crisis puts major institutions at risk

Sir — J. L. Cabrera *et al.* (*Nature* **421**, 786; 2003) and C. Mendoza *et al.* (*Nature* **421**, 473; 2003) present opposing views in their Correspondence letters about the amount that the Venezuelan government spends on promoting research and development. Although Cabrera *et al.* are correct to say this budget was higher last year than it was 10 years ago, their conclusions are based on inaccurate and misleading data.

A much better indicator is percentage of gross national product (GNP) dedicated to R&D. My analysis (see *Interciencia* **28**, 21–28; 2003) reveals a drop from 0.45% in 1990 to an all-time low of 0.26% in 2000, with a comeback to 0.43% in 2001. Going further back, history does not support the claim by Cabrera *et al.* of “continuous growth”, but reveals a pernicious cycle.

Indeed, since 1990, there have been no significant increases in the percentage of the national R&D budget assigned to most of the main research areas: on average, 7.5% to the Venezuelan Institute for Scientific Research (IVIC), 18.7% for the research facilities of public universities, 21.5% for the Ministry of Science and Technology, and a small amount for other institutes. Meanwhile, the percentage for

Intevep — the technology R&D centre of the state-owned oil company PDVSA — has grown from 30% to an average of 40% and rising: it received 55% in 2001.

Thus, the increase in R&D spending, as a percentage of GNP, to 0.43% — still not up to the 1990 level — is mainly due to increased funding for oil-industry research.

The comment by Cabrera *et al.* that “more than 60% of Venezuela's science budget comes from the government” suggests that the private sector may supply nearly 40%, but this is not the case. All the above organizations are publicly funded and constitute most of Venezuela's R&D.

However, I agree with Cabrera *et al.* that the current crisis will have a negative impact on this year's scientific activity and budget, because institutions have suffered cuts of 13–35%. Currency exchange controls will lead to the collapse of scientific institutions such as IVIC and its cherished library (*Nature* **421**, 682; 2003). And according to the newspaper *Ultimas Noticias* on 21 February 2003, 881 scientists and technologists from a workforce of 985 professionals at Intevep were laid off on 4 February, and its budget was slashed by three-quarters, leaving “only 10% dedicated to research”.

Jaime Requena

Academia de Ciencias Físicas, Matemáticas y Naturales, PO Box 80383, Caracas-1080A, Venezuela

If atomic precision is unfeasible, so is life

Sir — Your Editorial article “Nanotech is not so scary” (*Nature* **421**, 299; 2003) attributes to me the idea of building devices that replicate “by manipulating atoms one at a time”, and points out that several leading figures in nanotechnology research argue that this is unfeasible. As well they might. My proposal is, and has always been (see *Proc. Natl Acad. Sci. USA* **78**, 5275–5278; 1981), to build atomically precise structures by using molecular machinery to direct conventional chemical reaction events with sub-nanometre positional control.

If this is fundamentally unfeasible, then so is life. Thus, these critics are mistaking atomic precision for atom-by-atom manipulation, while failing to address the actual concepts analysed in the technical literature. These misdirected arguments have needlessly confused the discussion of genuine long-term safety concerns.

K. Eric Drexler

Foresight Institute, 123 Fremont Avenue,
Los Altos, California 94022, USA

correspondence

Contributions, preferably less than 500 words long, may be submitted to corres@nature.com.

The politics of publication

Authors, reviewers and editors must act to protect the quality of research.

Peter A. Lawrence

Listen. All over the world scientists are fretting. It is night in London and Deborah Dormouse is unable to sleep. She can't decide whether, after four weeks of anxious waiting, it would be counterproductive to call a *Nature* editor about her manuscript. In the sunlight in Sydney, Wayne Wombat is furious that his student's article was rejected by *Science* and is taking revenge on similar work he is reviewing for *Cell*. In San Diego, Melissa Mariposa reads that her article submitted to *Current Biology* will be reconsidered, but only if it is cut in half. Against her better judgement, she steels herself to throw out some key data and oversimplify the conclusions — her postdoc needs this journal on his CV or he will lose a point in the Spanish league, and that job in Madrid will go instead to Mar Maradona.

The decision about publication of a paper is the result of interaction between authors, editors and reviewers. Scientists are increasingly desperate to publish in a few top journals and are wasting time and energy manipulating their manuscripts and courting editors. As a result, the objective presentation of work, the accessibility of articles and the quality of research itself are being compromised.

One main cause

These trends are fuelled by the increasing pressure in biomedical science to publish in the leading journals. Even our language reflects this obsession — we say that Jim Jargon did well as a graduate student because he published a “*Cell* paper”, illustrating that we now consider the journal to be more important than the scientific message. If we publish in a top journal we have arrived, if we don't we haven't.

Why has this happened? It is partly because, rather than assessing the research itself, those who distribute the money and positions now evaluate scientists by performance indicators (it is much easier to tot up some figures than to think seriously about what a person has achieved). Managers are stealing power from scientists and building an accountability culture that “aims at ever more perfect administrative control of institutional and professional life”¹. The result is an “audit society”², in which each indicator is invested with a specious accuracy and becomes an end in itself.

Evaluations of scientists depend on numbers of papers, positions in lists of authors, and journals' impact factors. In



Strategic submissions: playing the game has become all-important.

Japan, Spain and elsewhere, such assessments have reached formulaic precision. But bureaucrats are not wholly responsible for these changes — we scientists have enthusiastically colluded. What began as someone else's measure has become our (own) goal. Although there are good reasons for publishing papers where they are more likely to be read, when we give the journal priority over the science we turn ourselves into philistines in our own world.

Some scientists realize this, but why have most taken up the journal chase so enthusiastically? It has to do with both psychology and careerism. Young researchers see a paper

in a good journal as their initiation into the scientific elite. The established seek publication in leading journals to certify their high opinion of themselves. All are learning that building capital in the hard currency of the audit society can be safer and easier than founding a reputation on discoveries. Another factor is that contemporary society has a craze for publicity, to which scientists are not immune. Many are gratified to find themselves or their work reported (accurately or not) in the media, and leading journals provide a route through press releases. *El País*, for example, usually features articles about any work by Spanish scientists published in *Nature*, *Cell* or *Science*.

Consequences

There are consequences for authors, editors and reviewers.

Authors have to decide when and how to write up their work. The ideal time is when a piece of research is finished and can carry a convincing message, but in reality it is often submitted at the earliest possible moment (two papers count for twice as much as one, never mind if the second paper mainly corrects errors in the first). Findings are sliced as thin as salami and submitted to different journals to produce more papers.

I don't need to tell you how things are, Miss Franklin. Non-scientists think of science as universal. Celestial, even. But science is terrestrial. Territorial. Political.

William Nicholson

Work must be rushed out to minimize the danger of being scooped — top journals will not consider a paper if a similar result has appeared in a competing journal, even if the experiments have taken years and there is only a week or two of disparity. Yet it can be advantageous if rival papers are submitted at the same time, as each author can use the other paper to tempt editors into concluding that the topic is a hot one. This practice has led to many dangerous liaisons between competing groups. It is no wonder that agonizing over presentation as well as the timing of submission keeps many scientists awake at night.

Authors need to decide how to get their paper into a top journal. Can the results be hyped to make them look more topical? Are there some trendy stock phrases that can be used³? Would oversimplification add to the appeal? Could a lofty take-home message be made to fit? Can even a tenuous link to a human disease be found? (Mention of a human disease boosts the number of subsequent references to the paper and can make it more attractive to a journal.) Can the results be squeezed into a shorter format than they require? For example, can they be submitted as a brief Letter to *Nature* — even though a longer paper in a more specialized journal would be of greater service to readers? Letters to *Nature* and Reports in *Science* are often presented in such a compressed form with such minuscule figures that they can be hard to decipher. Supplementary online material may alleviate this problem, although readers of print editions may not find it convenient to look at, and people have concerns about the length of its electronic shelf-life.

Increasingly, such a high premium is put on presentation that the leader of a group (who has not done the experiments) writes the paper reporting work done by a junior scientist (who has). The team leader is more experienced and more able to present the work in the best possible light — and for this, a lack of knowledge of the details can be advantageous! The student or postdoc is released to go back to the bench, increasing productivity. However, she or he does not get taught how to write up results⁴.

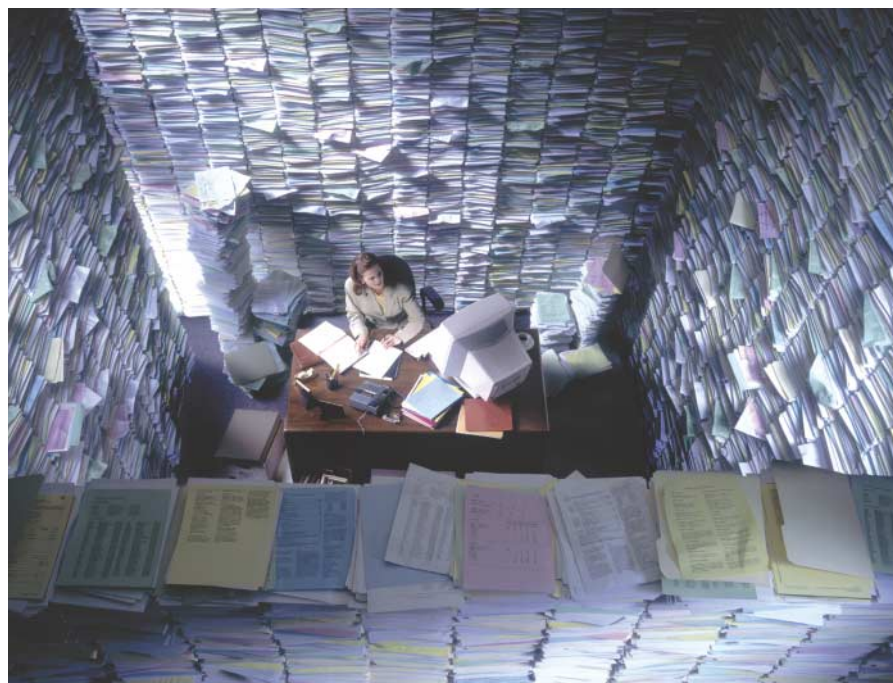
Editors. It is no surprise that editors of elite journals receive many submissions. For example, *Nature* now receives around 9,000 manuscripts a year (double that of 10 years ago) and has to reject about 95% of biomedical papers. *Development*, a quality specialist journal, now rejects roughly 70%, compared with 50% in 1990. In leading journals there are too many submissions to send most out for peer review, so the editor's decision has become, quantitatively, much more important than the judgement of reviewers. Consequently, editors are courted by authors who resort to tactics such as charm offensives during "presubmission enquiries", networking at conferences and wheedling telephone

calls — or pulling rank, using contacts, threatening and bullying. Group leaders can justify spending time and ingenuity on these stratagems — editors can be swayed and the rewards for success are high. Furthermore, impact factors and finance have joined forces to build up competition between top journals (Cell Press was recently sold for a great deal of money). One result is that editors are sent out to woo star scientists for their trendiest papers. These forces all combine to create an antiscientific culture in which pushiness and political skills are rewarded too much, and imaginative approaches, high-quality results and logical argument, too little.

Even experienced editors are on uncertain ground — sifting through a mass of diverse papers objectively and hurriedly is almost impossible. The advent of the Medline search and other Internet-based services has helped them, but it is still difficult to see clearly into the dark corners of specialization. Understandably insecure, editors play safe and favour the fashionable, familiar and expected over the flaky and unexpected — or original. Inevitably, mistakes are made. The original paper by Michael Berridge and Robin Irvine on phosphoinositol and signalling, which became the second most quoted article throughout the 1980s, was originally turned down by *Nature*. The authors fought back and it was accepted⁵. But when Berridge synthesized the information and added new ideas in another paper, it was rejected again by *Nature*, eventually published by the *Biochemical Journal*⁶ and became the fifth most quoted paper of the 1980s⁷.

Reviewers are, of course, authors wearing a different hat. There can be conflicts — for example, does the reviewer favour the work of a competitor and thereby endanger his or her student's career? Such opposing interests can explain why two reviewers of similar expertise sometimes present vastly different opinions about the same paper. It does not help that top journals are increasingly giving reviewers an extra task. Apart from the traditional technical and scientific assessments, where objective criteria are paramount, reviewers are now being asked to judge whether a manuscript constitutes a "Science" paper — is it sufficiently exciting to interest the "general reader"? This participation in editorial decisions gives reviewers opportunities to punish authors they do not like, settle old scores and hold up competitors. From many years of editing experience, I am persuaded that a minority of reviewers take advantage of these opportunities. Some bounce the same paper from more than one journal, making it more difficult for a less politically adept scientist to present his or her work, especially if it goes against the current grain. Objectivity is also threatened by a tacit understanding between some leading scientists: they invite each other onto committees, to conferences, nominate each other for prizes and awards, and support publication of each other's papers.

Another relatively recent phenomenon is the practice of sending papers to three reviewers. Although this is partly to ensure that at least two reviews are received, I think it is also so that the advice received cannot be a tie. Decision by vote can encourage rejected authors to make empty appeals, praising favourable reviewers, denigrating negative



Après le déluge: searching for the needle of quality in the haystack of submissions.

GETTY IMAGES

ones and asking for new reviewers — in the hope of getting another plus. Rejection is easier to accept if there is a thoughtful reason for it from which one can learn.

Hard-pressed editors take power from authors and hand it to reviewers in other ways. Reviewers often ask for changes and new experiments, even though they may be rather ignorant of the details and may have formed an opinion of the paper in half an hour. Nevertheless, the easiest and most commonly chosen course for the editor is to ask the authors to “satisfy” all of the reviewers, then send the revised manuscript for reassessment. If authors have well-founded disagreements with a reviewer, they find themselves in a dilemma: do they invest time in experiments they do not believe will help, do controls that few other informed people would find important, or even draw conclusions that are not theirs? If they do not, reviewer X may not be appeased and the editor would be unswayable. In former days, these authors could have solved their problems by sending their papers elsewhere, but now that the journal itself has become so important to their careers, they feel forced to comply. In this situation the reviewers can become more like censors than assessors. I have seen many examples of this and, sometimes, months of research time have been misspent, even to the extent that an author can be scooped in the interim.

Faster publication times, materials-transfer agreements and threats of legal action to force journals to identify reviewers have added to the pressure. In the case of faster publication times, journals can offer chosen authors fast-track treatment and advanced online publication, helping them to steal a march on competitors. A reviewer can use information and may have time to modify his or her own manuscript and even publish it elsewhere first. Temptation and suspicion have heated up enough to melt the wall of confidentiality that reviewers owe to authors. Still, I believe there is genuine confusion about the level of confidentiality that reviewers should adopt. Is the reviewer obliged not to reveal even the existence of a submitted manuscript to anyone? I think so, but do we all concur? Should a reviewer agree to assess a paper that he or she has already advised another journal to reject? I think not, but this happens frequently.

Cures?

It is no wonder that authors are becoming paranoid. Roughly half of the submission letters I now receive request me not to use certain reviewers, often because of “conflicts of interest”. Behind this phrase lurks the fear of misuse of the information in the paper — although admittedly it is sometimes a ploy to avoid the sharp-eyed and critical.

My main purpose here is consciousness-raising. But we can all start to improve things

There is no form of prose more difficult to understand and more tedious to read than the average scientific paper. **Francis Crick**

by toning down our obsession with the journal. The most effective change by far would be if the organizations that award grants and manage research programmes were to place much less trust in a quantitative audit that reeks of false precision. Such organizations have the big advantage of hindsight — unlike editors and reviewers at the time of submission, they can ask themselves if key papers published by the candidate are illuminating, have proved influential and whether their main results have been confirmed by others.

Authors can help to break up the cult of the journal. One way is to set up mutually supportive alliances, as has been done for the field of cell signalling (<http://www.signaling-gateway.org>). If established authors start to publish selectively in open-access websites and in specialized journals when appropriate, a better example would be set for younger scientists. This would reduce the enormous pressure on the leading journals, which then could again begin to publish more comprehensible papers that tell a complete story, perhaps even bringing the ‘general reader’ back to life.

I am not suggesting sweeping changes to the review process. For example, I don’t think a change to open peer-reviewing (as discussed in ref. 8) will help, mainly because younger reviewers would be intimidated and the political power of the established would be increased. One change which would now be feasible through online submission of two forms of the manuscript, would be to deny the reviewers authors’ names. It is crucial that the responsibilities and duties of reviewers are clarified and made more public. For example, they could be better educated about confidentiality by the journal, along the lines of *Nature’s* advice (<http://www.nature.com/nature/submit/policies/index.html#8>).

Professional editors need to be more aware of these dangers. They now have to make difficult decisions that are of vital importance to authors, far beyond the publication or not of a particular paper, as well as meeting rejection rates as high as 95%. They have — perhaps understandably — been relinquishing too many of their responsibilities to reviewers. It does not help that editors may not have had enough experience of research and lack hands-on knowl-

edge, particularly outside one narrow subject area. They need to act now to reinstate authors’ rights. Once a decision has been made to publish in principle, they should never simply demand in a blanket sense that authors satisfy reviewers X, Y and Z, but should interpret referees’ advice and be willing to accept reasoned discussion about aspects of the referees’ criticisms. Editors should then be in a decision to adjudicate among themselves or to seek further opinion from an expert who is given both sides of the argument. Editors should appreciate that, unlike the authors whose names are out there, anonymous reviewers will not be held to account if they make a mistake. It should always be remembered that the proper role of the reviewer is to advise the editor, not to gain control over the author’s paper.

Editors should also take a more long-term and broader view about what is of interest, and act positively to encourage new approaches and topics in an affirmative action against fashion. It is fashion that makes looking for new members of signalling pathways into the hottest of current topics, which can lead to unnecessary duplication. Just one example — no less than four independent studies on the same new gene (*pygopus*), each describing years of careful and hard work by several people, have just been published (see ref. 9 and references therein).

As authors, we have abandoned the attempt to make our experimental papers accessible or comprehensible to the nonspecialist, often writing undiluted mixtures of hype and jargon. This is partly because we are writing in shorthand to fit our papers into a small space, and partly because we are trying to con the editors. But why not write papers that are readable, reduce the number of acronyms and gobbledeegook, and put methodological details in supplementary material on the web?

It is we older, well-established scientists who have to act to change things. We should make these points on committees for grants and jobs, and should not be so desperate to push our papers into the leading journals. We cannot expect younger scientists to endanger their future by making sacrifices for the common good, at least not before we do. ■

Peter Lawrence is at the MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK. e-mail: pal@mrc-lmb.cam.ac.uk. He has edited the journal Development since 1976, served on the editorial board of Cell and EMBO J., authored many papers and reviewed many more.

1. O’Neill, O. *A Question of Trust* (Cambridge Univ. Press, 2002).
2. Power, M. *The Audit Society: Rituals of Verification* (Oxford Univ. Press, 1997).
3. Lawrence, P. A. *Nature Rev. Genetics* **2**, 139–142 (2001).
4. Lawrence, P. A. *Nature* **415**, 835–836 (2000).
5. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315–321 (1984).
6. Berridge, M. J. *Biochem. J.* **220**, 345–360 (1984).
7. Pendlebury, D. *The Scientist* **4**, 18–21 (1990).
8. Gura, T. *Nature* **416**, 258–260 (2002).
9. Belenkaya, T. et al. *Development* **129**, 4089–4101 (2002).

Counting the cost

Can cost–benefit analysis solve the problem of assessing environmental risk?

Risk and Reason: Safety, Law, and the Environment

by Cass R. Sunstein

Cambridge University Press: 2002. 352 pp.

£25, \$30

Raymond T. Pierrehumbert

How much would you pay to save an isopod? What's that, you say, they're slimy little devils that you'd rather see exterminated? Keep your pencils sharp, because your answers to this and a few thousand other questions (how much would we have to pay you to let us expose your children to malaria?) will be used to determine whether it's worthwhile trying to halt global warming. Oh, and by the way, please let us know how you think the next hundred generations of your descendants will value these things, too.

Risk and Reason — a rather sprawling book by legal scholar Cass Sunstein — covers a lot of interesting territory pertinent to environmental regulation. The main theses are that people can't be trusted to judge rationally the risks they face, and that the answer to this problem is a priesthood of government technocrats who carry out cost–benefit analysis (CBA) insulated from judicial review and pressure by advocacy groups. Unlike those who declare that things are great and getting better all the time — a view espoused by the likes of Bjørn Lomborg, Gregg Easterbrook and Aaron Wildavsky — Sunstein is no ecopollyanna. He admits the reality of environmental problems and, while disparaging the methods of what he calls "70s environmentalism", he copiously celebrates its accomplishments. His stated goal is to get people to "listen to their values, not their mistakes", and I have no reason to doubt the sincerity of this intent. My fear is that Sunstein's cost–benefit state would succeed only in getting people to listen to the mistakes of their accountants instead.

Much of *Risk and Reason* is devoted to analysing ways in which the public has overreacted to supposedly small risks. Sunstein rather too readily, in my view, buys into the premise that toxic sludge is (relatively) harmless. But let us concede that the means by which society allocates resources to various risks is flawed. Is CBA a useful tool for improving the situation? The difficulty faced by CBA is that the complete description of the effect of an action consists of an exceedingly long list of characteristics that generally share no common yardstick for quantitative comparison with each other. These include such things as lives lost, disease incidence, species extinction, habitat loss and loss of freedoms, as well as uncertainties in estimates and the distribution of



Saving the world: the Whole Earth Jamboree typified the environmental movement of the 1970s.

effects among the millions of affected parties.

Economists often like to collapse this enormous space onto a one-dimensional scale called 'utility' (money, to the rest of us). Trying to understand cost–benefit trade-offs in this way is like trying to understand the structure of the Eiffel Tower by squashing it flat and rolling it into a wire. Reduction to utility destroys the information that people need to make value judgements, and leads to such absurdities as an early Intergovernmental Panel on Climate Change (IPCC) report that valued the life of a New Yorker as equal to 15 African lives. The discounting required to deal with the future in CBA is also problematic: at a 3% discount rate, it is better to save 100 lives this year than to avoid the extinction of the entire human race in 650 years. Sunstein is aware of such problems, but sees their solution as mere technicalities to be dealt with by the appropriate agencies.

In contrast, Amartya Sen in *Rationality and Freedom* (Harvard University Press, 2002) concludes that the sort of CBA that Sunstein mostly advocates is "not so much a discipline as a dream". But Sen offers much wisdom about the prospects for construing costs and benefits more broadly. Environmental justice can also be founded on a theory of rights, such as for free speech. Peter Singer's perceptive essays in *One World* (Yale University Press, 2002) provide an invigorating discussion of this perspective, which is encountered only briefly in *Risk and Reason*.

Sunstein says that CBA does not necessarily stack the decks in favour of industry

groups seeking to obstruct regulations, but he provides scant reason for such optimism. He cites the Montreal Protocol as an example of a treaty that was stimulated by CBA, but the history of this treaty reveals little use of the CBA but much use of the familiar methods of '70s environmentalism. And the case law reviewed here is all anti-regulatory. Indeed, the efforts of the present US administration to strengthen the role of CBA are hardly encouraging, particularly the tendency to fill advisory panels with members who will provide ideologically reliable advice. Having seen some "2000s environmentalism" in action, the '70s type looks distinctly attractive to me.

Surely it is worthwhile to pursue cost-effectiveness in achieving an agreed goal, to make better use of sound science in policy, and to allow more open discussion of costs as one factor among many — none of which require CBA. *Risk and Reason* does reveal all this, but only after you've mastered the fortitude to look past its overarching technocratic vision. In a democracy, there is no substitute for a plurality of well-informed voices, something that, at times, seems to make Sunstein distinctly nervous. Interest groups from Greenpeace to the American Petroleum Institute, and trusted authorities such as the IPCC, all have their place in the chorus. But there will be little hope of the public making sense of the cacophony without efforts to address the appalling state of science education. ■

Raymond T. Pierrehumbert is in the Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA.

The smart bird of the south

The Adélie Penguin: Bellwether of Climate Change

by David G. Ainley

Columbia University Press: 2002. 314 pp.

\$59.50, £42.50

Yvon Le Maho and David Grémillet

Penguins fascinate humans. They appeal to kids and adults alike, they hop around on television and their images fill toy stores. They are charismatic creatures par excellence. But penguins also intrigue scientists because surprisingly little is known about these seabirds, especially about their lives far offshore. Take the Adélie penguin. This 'little man in evening dress' only became known to science on 21 January 1840, when the French explorer Jules-Sébastien-César Dumont D'Urville and his crew reached the Antarctic mainland. D'Urville's men landed on some rocks off what he named 'Terre Adélie' after his wife Adèle. A painting at the National Maritime Museum in Paris recalls this event, showing how the crew of the *Astrolabe* hurtled down an island, carrying away some penguins, which were later described as the inhabitants of the newly discovered land.

Over 160 years later, David Ainley's book reminds us that the Adélie penguin is not a trophy, nor a pleasant mascot, but a most astonishing bird and a key element of the fragile Antarctic ecosystem. Ainley has been researching Antarctic and Pacific seabirds for some 30 years, so he is certainly the right person to compile this book. He was one of the early pioneers of the idea of Antarctic food webs linking penguin population dynamics with large-scale climatic features. Ainley shows that Adélie penguins exploit



Voyage to a new land: this painting shows the discovery of the Adélie penguin on Terre Adélie in 1840.

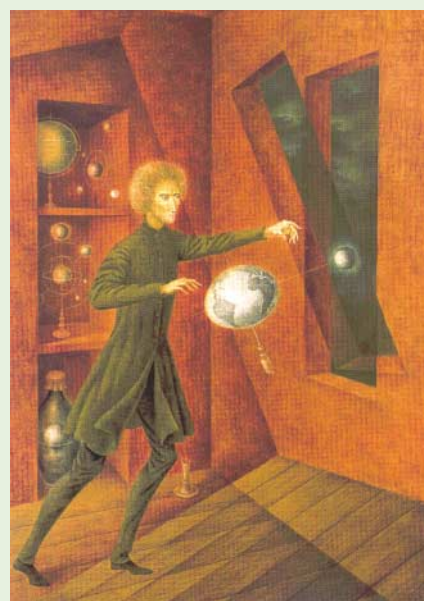
the pack-ice zone better than any other seabird species, and that the extent of sea-ice cover around Antarctica conditions the lives of the 2.5 million Adélie penguins. He articulates his book around this idea, and this attractive volume addresses all of the major aspects of the bird's biology. It will be a marvellous source of information and inspiration for generations to come.

Ainley has been highly involved in a long-term study of the breeding biology of Adélie penguins at Cape Crozier in the Ross Sea. He presents the data set in detail, focusing on the breeding biology of birds from this area. Other crucial aspects, such as the bird's feeding behaviour, are unfortunately only briefly discussed. In that sense *The Adélie Penguin*

seems to comprise two different books. One is a detailed monograph on the breeding biology of the Adélie penguin in one particular location; the second is a brief yet attractive summary of the penguin's ecology in a global context. This leaves the reader to wonder whether the catchy subtitle was the author's choice, or rather reflects the marketing strategy of the publisher. In our view, Ainley's global perspective of Antarctic ecosystems would have been enriched by studies at more field sites, perhaps through collaborations outside the US community.

Beyond these minor shortcomings, Ainley consistently shows a rigorous, scientific approach. He is rightly concerned about the impact of scientific methods on the well-being of the animals. He was among the first to ask fundamental, pertinent questions about the impact of penguin ringing, for example. He was right from the start: penguin ringing might indeed reduce survival chances or breeding success, although the nature and the extent of the effect are still to be clarified.

Ainley had the tenacity and patience to run a long-term programme at Cape Crozier, gathering field data, year after year, in particularly difficult times when it was unfashionable to focus on anything other than molecular biology and biotechnology. Nowadays, links between the behaviour of marine top predators, the mechanics of marine food chains and global climate patterns are more than obvious. Suddenly, long-term studies of seabirds are sexy. As Ainley emphasizes in his book, scientists are now building rapidly on these immensely valuable background data. Using miniaturized electronic devices attached to the birds,



A relative view of the Universe

Spanish surrealist painter Remedios Varo created the metaphor shown here for Einstein's space-time Universe in 1963.

The Phenomenon of Weightlessness, as it is called, is one of over 300 illustrations in *Exploring the Invisible: Art, Science and the Spiritual* (Princeton University Press, \$49.95) by the art historian Lynn Gamwell. The book, which is scholarly but highly readable, follows the history of the way that scientific thought and discovery has influenced art, particularly since the latter half of the nineteenth century.

they can explore the marine lives of Adélie penguins (these birds spend less than 10% of their lives on land) and further improve our understanding of oceanic ecosystems. ■

Yvon Le Maho and David Grémillet are at the Centre d'Ecologie et Physiologie Energétiques, UPR CNRS, 23 rue Becquerel, 67087 Strasbourg Cedex 2, France.

Networks untangled

Six Degrees: The Science of a Connected Age

by Duncan J. Watts

W. W. Norton/Heinemann: 2003. 448 pp.

\$27.95/£17.99

Lada Adamic

Advances in computing over the past decade have not only improved the world's networks, they have made it easier for scientists to study networks, both new (the Internet) and old (metabolic and other biological networks that have been evolving for billions of years). The recent avalanche of research on networks was set off by Duncan Watts, who, while working on his PhD thesis with Steve Strogatz at Cornell University, wrote the seminal article "Collective dynamics of 'small-world' networks" (*Nature* **393**, 440–442; 1998). Five years and hundreds of articles later, the research community is making progress in understanding the structure of networks. But working out how network structure affects dynamics, which was Watts' original interest, has proven to be a much harder problem. Undeterred, Watts has continued this vein of research, with several interesting and rewarding results. This book is a narrative of his research and of the many connections that have led him down this path.

New in paperback

Megawatts + Megatons: The Future of Nuclear Power and Nuclear Weapons

by Richard Garwin and Georges Charpak

University of Chicago Press, \$20

"We now have a book in which all aspects of the interaction between these two applications of nuclear energy are acknowledged and frankly examined, and the awesome global implications laid bare." Stuart Young, *Nature* **414**, 487–488 (2001).

Our Cosmic Habitat

by Martin Rees

Phoenix Press, £7.99

A Thin Cosmic Rain

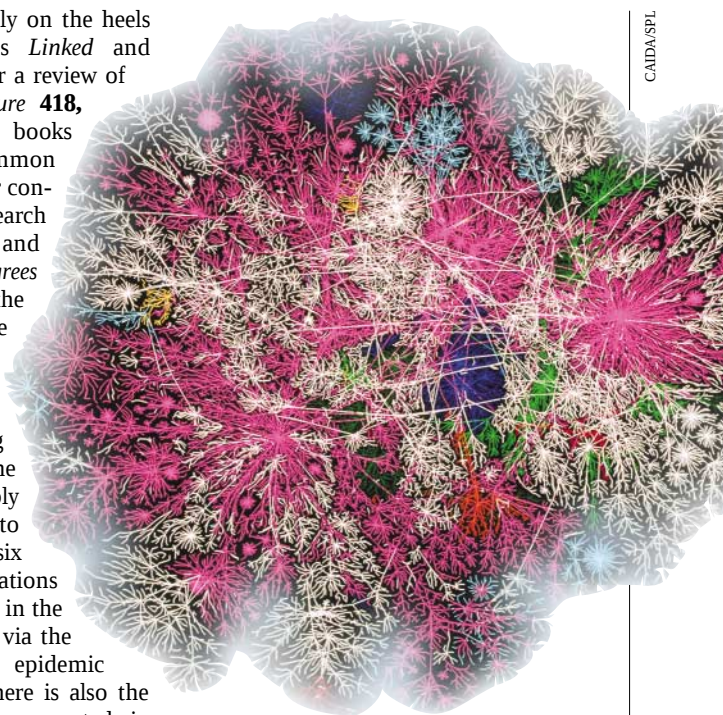
by Michael W. Friedlander

Harvard University Press, \$17.95, £11.95, £18

Six Degrees follows closely on the heels of Albert-László Barabási's *Linked* and Mark Buchanan's *Nexus* (for a review of these two books, see *Nature* **418**, 127–128; 2002). All three books cover quite a bit of common ground, including the major contributions to network research made by both Watts and Barabási. In addition, *Six Degrees* and *Linked* use many of the same examples to illustrate the role of networks. There is the groundbreaking experiment in the 1960s by Stanley Milgram showing that any two people in the United States (and probably the world) are connected to each other through a mere 'six degrees'. There are the observations that networks are important in the spread of computer viruses via the Internet, and the AIDS epidemic through sexual contacts. There is also the certainty that everything is connected: in 1996, the failure of a single transmission line in the US electrical grid caused major power outages across the western United States; and the 1997 currency-exchange crisis in Thailand was felt not only in Southeast Asia, but worldwide.

Six Degrees and *Linked* differ in their focus, however. Barabási tells a simple and convincing story: that networks in many systems arise through a rich-get-richer phenomenon, yielding many poorly connected nodes and a few hubs, or extremely well-connected nodes. These hubs affect properties of the network, such as susceptibility to computer-virus epidemics. Watts' book presents a broader view: that the distribution of connections among the nodes is just one of the factors that influence the dynamics of networks. In the context of social networks, some individuals may be better connected than others, but who we know also depends on who our friends know, and where we live and work, and this leads to a network property called 'clustering'. Watts' original *Nature* article showed that networks can be clustered, but as long as there are a few random connections, the network is a 'small world', with only a few hops separating any two nodes.

But the existence of short paths does not, in itself, explain how people can find their targets. If you are trying to reach someone, but know only your friends, and some of your friend's friends, six degrees can seem quite distant. Watts has developed a mathematical model to show why individuals succeed in finding short paths simply by choosing for each link in the chain a friend who is most like the target. In addition to presenting a solution to the small-world problem, Watts has studied information cascades and the spread of disease in networks. The insights



An understanding of networks can show us how computer viruses spread on the Internet.

presented, though quite general, are based on the results of analysis and simulation, and would benefit from empirical validation.

Watts' casual style of writing, dotted with humorous introspection and observations about his collaborators, is accessible and engaging. The more involved models covered in the later chapters require a bit of effort to follow, even in the absence of equations. For further reading, Watts provides a list of resources, grouped by topic.

Six Degrees primarily covers Watts' own work, so it is not a comprehensive overview of recent developments in the field (the books by Buchanan and Barabási are somewhat broader). But it does provide a good introduction to the topic, including many interesting real-world examples of network dynamics. It also offers historical background and several new results relevant to the social sciences. If you haven't yet had time to learn about the latest intriguing research on networks, reading this book could help you see why people increasingly believe that understanding networks is the key to such seemingly disparate problems as securing the Internet, fighting epidemics, curbing terrorism and deciphering genetics. ■

Lada Adamic is at HP Labs, 1501 Page Mill Road, ms 1139, Palo Alto, California 94304-1126, USA.

Correction In his review of Cassell's *Laws of Nature* by James Trefil (*Nature* **421**, 578–579, 2003), Walter Gratzer imputed to the author an error that he did not make: it is blue-eyed parents who do not have brown-eyed children, brown being dominant, and not the other way around. We apologise to James Trefil.

A comic-book hero

Summarize yourself in the form of a title of a paper in Nature.

Rapid degeneration of an experimental biophysicist into an ad agent through competitive inhibition by young colleagues.

What was your first experiment as a child?

On a sunny day, my younger sister and I were strolling on the campus of the Japanese Imperial College. I had stolen a box of matches, with which we began to investigate if mechanical energy could trigger a chemical reaction. Minutes later, my sister pointed at a place a few metres away, where I saw a belt of black moving steadily away from us, leaving shrunk and curly grass behind. An adult came. The two kids walked away calmly, talking slightly too loudly about how they knew nothing. The lesson: detection of single fluorophores requires darkness.

Who has been the most important mentor in your career?

My father, in the most indirect way. I can count the number of words we have exchanged since I was born. Of those words, I still remember: "Geocentricism and heliocentricism are equivalent, except that the physics is simpler with the latter."

What makes a good scientific mentor?

I wish I knew.

Whose graduate student would you most like to have been?

Einstein, if I could rewire my brain. Peter Mitchell, to reconfirm that I can never be creative.

What single scientific paper or talk changed your career path?

Hugh Huxley's electron micrograph showing the sliding mechanism of muscle contraction. It was the first time in my life that I saw an 'explanation' in biology, the branch of science I had disliked most until then.

What book has been most influential in your scientific career?

Lubert Stryer's *Biochemistry*, a textbook that even a physicist can appreciate. It also teaches us how to write clearly: where Stryer fails to be clear, researchers in the field are to blame.

What gives you the most job satisfaction now? What are your major frustrations?

All I can do now is envy the (infrequent) achievements of my young colleagues made without my aid. Only bench work really satisfies me, but I no longer have the energy to sit down at a bench for more than five minutes.

What literary character would you employ as a postdoc?

Imaginary postdocs? No thanks.

What's your favourite conference destination, and why?

Aspen, for the obvious reason.

What book is currently on your bedside table?

The Tale of Genji, a 1,000-year-old novel by Murasaki Shikibu, recommended by a recent mentor of mine who teaches me thermodynamics. It is about noble women and men; the first two-thirds concerns the son of an emperor and the women he apparently loved. The ancient world is so different that I find no place for myself.

What music heads the playlist in your car or lab?

The soundtrack to *Flashdance*.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

If the food is excellent, anyone who speaks Japanese, or none. Otherwise, I would beg the company of Audrey Hepburn to see if she was as attractive in real life, too.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

I can only try to sleep.

What one thing would you rescue from your burning laboratory?

My laptop or human life? That is the question.

What's the best piece of advice you've ever received?

That a grant reviewer might appreciate your proposal if he/she feels that you are attempting something beyond your ordinary power.

Where and when would you most like to have lived or worked?

I would certainly enjoy staying in present-day Japan, now that dining has become less expensive. For work, a university on either the west or east coast of a United States that has no intention of bombarding others.

What do you most dislike about having research published?

Until I became a postdoc, I believed that doing research to publish was sin, just as artists should not paint for money.

Name one extravagance you can now get away with because of your eminence.

To excuse myself from a meeting for another (possible) meeting.



Kazuhiko Kinosita Jr

Kazuhiko Kinosita Jr is a professor of single-molecule physiology at Okazaki National Research Institutes in central Japan. He likes skiing, mountain walking, comic strips, sarcasm and irony.

What single discovery, invention or innovation would most improve your life?

A garbage annihilator. I live a single life in a small Japanese city, where every family must sort its rubbish into more than ten categories. Once sorted — and usually cleaned and packed — the various rubbish types must be taken to different stations. These are often a long way from each other, and tend to be open until 8:30 a.m. just one day a fortnight (kitchen waste can be taken to a station as often as twice a week on specified mornings). Newspapers, for example, must be tied with string, not bagged, and taken to a city hall kilometres away from where I live. I have heard of a university professor in a nearby city who resigned and moved to Tokyo (where life is less stringent) after he found himself cutting newspapers into pieces every night.

What do you do to relax?

I read comic strips. Actually it's not that relaxing because I am so absorbed by them.

What would you have become, if not a scientist?

A computer programmer. I would have tried to invent a new operating system.

What music would you have played at your funeral?

I have already told my last wish to my wife — that my death is to remain undisclosed for two months.

How would you like to be remembered?

An offensive guy, but not totally intolerable.

What's just around the corner?

Another rejection letter from *Nature*. ■

Just mix and patch

Keith Mostov and Mirjam Zegers

A single layer of epithelial cells makes up the primary barrier to infection in our bodies. This layer is poised to repair damage quickly, patching up any holes to keep invading organisms out.

More than 90% of infectious micro-organisms, including HIV, enter the body through exposed mucosal surfaces such as those in the digestive and respiratory systems¹. These surfaces are lined by a single layer of cells, known as epithelial cells. This delicate layer is easily damaged, allowing invaders and toxic substances in — so it is essential that any holes are filled rapidly². On page 322 of this issue, Vermeer and colleagues³ reveal a mechanism by which epithelial layers stand ready to begin repairs within moments of injury.

A defining feature of epithelial cells is that they are polarized (spatially asymmetric)⁴. All cells are bounded by a plasma membrane composed of a protein-rich lipid bilayer. In epithelial cells this membrane is divided into two domains: the apical domain faces the central space of the organ, and the basolateral domain faces adjacent cells and underlying tissue. These domains perform different jobs and have different compositions. In addition, the apical and basolateral domains are separated by specialized junctions between neighbouring cells, known as tight junctions, which surround the cells like rubber o-rings⁵. Tight junctions keep proteins in one domain from diffusing into the other, and also act as a seal between adjacent cells to prevent large molecules from crossing the epithelial layer.

When an epithelial monolayer is damaged, for instance by mechanical injury or toxic substances, the tight junctions can no longer carry out these functions. Vermeer *et al.*³ have discovered how this loss of tight-junction integrity allows epithelial cells to sense their injury and launch a repair programme that promotes cell division to replace cells that are lost (Fig. 1).

At the heart of the matter is the ability of epithelial cells to activate a protein, erbB2, which is a receptor in the plasma membrane that promotes cell proliferation⁶. When a specific protein — a ligand — binds to the portion of erbB2 that is exposed on the outside of the cell, erbB2 sends a signal to its intracellular portion. That intracellular portion is a tyrosine kinase, an enzyme that transfers a phosphate group from ATP molecules to tyrosine amino acids in other proteins. The phosphorylated tyrosines then act as binding sites that recruit many more proteins to the inner surface of the plasma membrane, and these proteins in turn

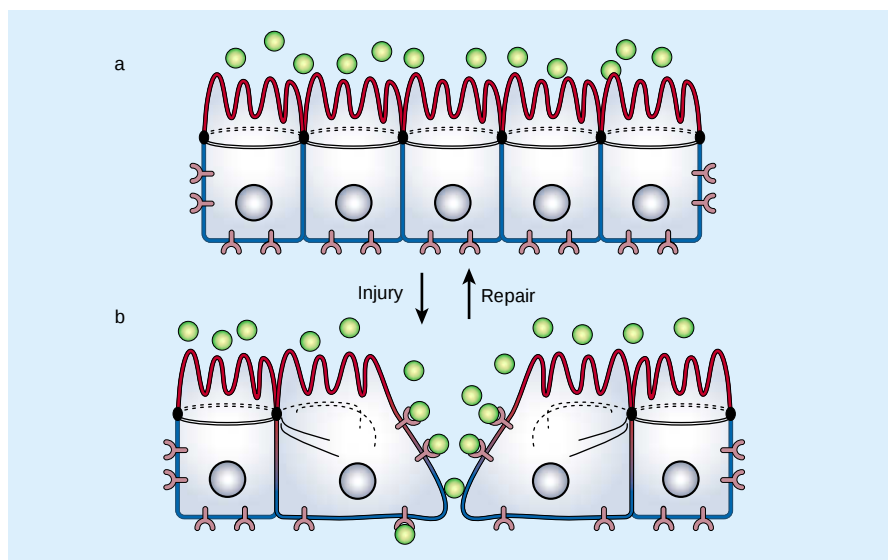


Figure 1 Tissue, heal thyself. Epithelial cells have an apical surface (red), which is separated by tight junctions (black) from the basolateral surface (blue). a, Vermeer *et al.*³ have found that the receptor erbB2 (purple) is expressed on the basolateral surface of airway epithelial cells, and is physically segregated from its ligand, heregulin (green balls), which is also produced by these cells. b, Injury allows heregulin to interact with erbB2, promoting cell proliferation to repair the epithelial layer.

activate several signalling pathways that stimulate cell division.

Vermeer *et al.* now show that erbB2 is present on the basolateral domains of human epithelial cells lining the air passages of the lungs. Not only that, but the airway epithelial cells also produce a ligand for this receptor, called heregulin⁶, and so have the potential to stimulate their own proliferation.

Ordinarily, proliferation must be strictly controlled, lest cells multiply uncontrollably to form a tumour. Vermeer *et al.* have discovered how heregulin is normally prevented from activating the erbB2 located on the same cell. The trick lies in the polarization of epithelial cells. As the authors show, heregulin is normally anchored to the outer surface of the apical plasma membrane. In normal conditions, some heregulin is cleaved to release an active fragment that is found in fluid coating the airways. Functional tight junctions normally prevent heregulin or its active fragment from reaching erbB2 on the basolateral surface. But mechanical damage (which the authors caused by scraping cells off a monolayer *in vitro*) allows molecules from the apical and basal surfaces to mix: so heregulin can reach and bind to erbB2, activate its kinase activity and promote

proliferation. Vermeer *et al.* confirm the importance of this mechanism by blocking either heregulin or erbB2 with specific antibodies; healing of the epithelial monolayer is impaired.

Extracellular calcium ions are needed for cells to adhere tightly to their neighbours, and tight junctions stop functioning if extracellular calcium is removed. Vermeer *et al.* find that this, too, allows heregulin to bind to erbB2. Calcium removal may mimic the diffuse epithelial injury that is associated with certain diseases, such as smoking-related bronchitis.

This mechanism, which allows epithelial cells to stand poised to promote their own healing, is conceptually so simple that one wonders why it was not discovered before. It also has implications for development and cancer. The distinguishing feature of multicellular animals is that their cells are organized into tissues, the most fundamental type being epithelial tissues⁷. Indeed, the simplest multicellular animals — such as hydra — consist of two concentric cylinders of epithelial cells, with non-epithelial cells in between. During development, cells form tissues by adhering to and sensing their neighbours, using intercellular junctions

that include tight junctions. Such junctions not only provide adhesive and barrier functions, but are also essential to the control of polarity, differentiation and proliferation⁸. Perturbation of junctions and polarity can cause epithelial cells to proliferate wildly, leading to a multilayered architecture and, potentially, cancer. Indeed, some 90% of fatal malignancies in adult humans arise from epithelia.

We are only just beginning to understand how junctions control proliferation⁹, but Vermeer *et al.* have revealed a possible new mechanism. This type of process might be widely used in development to prompt epithelial cells to cease proliferating when they polarize and form junctions. It could also be a means by which the loss of both junctions and polarity causes uncontrolled cell proliferation in cancer.

Finally, these findings illustrate a broader concept: the spatial organization of signalling. Cells contain a huge number of possible signalling pathways, but these are often dormant. Signalling might be controlled by compartmentalizing signalling components.

For instance, it is known that signals can be transmitted from one side of the cell to the other by moving across the cell interior¹⁰. The novelty of the erbB2–heregulin system lies in the fact that, here, a signal is transmitted extracellularly around the surface of the cell. ■

Keith Mostov and Mirjam Zegers are in the Departments of Anatomy, and Biochemistry and Biophysics, University of California, San Francisco, California 94143-2140, USA.
e-mail: mostov@itsa.ucsf.edu

1. Bomsel, M. & Alfsen, A. *Nature Rev. Mol. Cell Biol.* **4**, 57–68 (2003).
2. Jacinto, A., Martinez-Arias, A. & Martin, P. *Nature Cell Biol.* **3**, E117–E123 (2001).
3. Vermeer, P. D. *et al.* *Nature* **422**, 322–326 (2003).
4. Nelson, W. J. & Yeaman, C. *Trends Cell Biol.* **11**, 483–486 (2001).
5. Mitic, L. L. & Anderson, J. M. *Annu. Rev. Physiol.* **60**, 121–142 (1998).
6. Yarden, Y. & Sliwkowski, M. X. *Nature Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
7. O'Brien, L. E., Zegers, M. M. P. & Mostov, K. E. *Nature Rev. Mol. Cell Biol.* **3**, 531–537 (2002).
8. Zahraoui, A., Louvard, D. & Galli, T. J. *Cell Biol.* **151**, F31–F36 (2000).
9. Balda, M. S., Garrett, M. D. & Matter, K. J. *Cell Biol.* **160**, 423–432 (2003).
10. Rojas, R. & Apodaca, G. *Nature Rev. Mol. Cell Biol.* **3**, 944–956 (2002).

Astronomy

Glowing embers

Tsvi Piran

The combined efforts of telescopes around the world have tracked the development of the 'afterglow' that follows a γ -ray burst, from only minutes after the burst until weeks later.

Gamma-ray bursts (GRBs) are short, intense flashes of γ -rays that arrive from random directions in the sky. They are followed by an 'afterglow' — decaying signals at longer wavelengths that may last weeks, months or even years. Until now, afterglows had only been picked up some time after the GRB, but, on page 284 of this issue, Fox *et al.*¹ report the continuous detection of an optical afterglow, beginning at almost the same instant as the burst itself.

Using data from optical telescopes in Japan and the United States, the observations close the gap that had existed between observations of the burst and observations of the afterglow. Combined with a later, worldwide observational campaign (involving 33 telescopes), these data provide unique information about the variation of the optical emission over time — the first, practically complete, optical light curve of a GRB afterglow. Fox *et al.*¹ also map out the afterglow of this GRB (GRB 021004) at X-ray wavelengths, using data from the space-borne Chandra X-ray Observatory. The results are exciting, but perplexing, because the light curves differ in several ways from what had been expected.

In the early 1990s, the Burst and Transient

Source Experiment (BATSE) on board NASA's Compton Gamma Ray Observatory revealed that GRBs originate at cosmological distances, far beyond the reaches of our own Galaxy. This means that GRBs are the brightest objects in the sky. BATSE provided a wealth of information on several thousand GRBs, but couldn't locate them accurately enough to facilitate a search for optical counterparts.

On 28 February 1997, however, the Wide Field Camera aboard the Italian–Dutch satellite BeppoSAX pinpointed the location of GRB 970228. Several hours later, BeppoSAX directed its Narrow Field X-ray camera towards this position and discovered an X-ray afterglow; discoveries of optical, infrared and radio afterglows followed, from which the GRB's host galaxy could be identified and a wealth of other information deciphered.

BeppoSAX transmitted data to Earth only once per orbit. Hence there was an inevitable delay between the detection of the burst and the observation of its afterglow. This gap in the data is critical, as some of the most interesting physical phenomena are expected to take place in this time window. In January 1999, the Robotic Optical Transient Search Explorer (ROTSE) received a triggering signal

from BATSE and turned towards the rough position of GRB 990123. ROTSE's very wide field of view enabled it to detect a prompt optical flash that coincided with the burst². But ROTSE's six snapshots covered only the first few minutes of the afterglow, leaving a large gap until other telescopes could swing into action several hours later.

The High Energy Transient Explorer II (HETE II)³ was built with the intention of bridging this gap. HETE II is capable of transmitting accurate GRB positions to the Earth in real time, to be followed up with immediate (automated or manual) optical observations. On 4 October 2002, 49 seconds after a GRB trigger announcing GRB 021004 and while the 100-second burst was still active, HETE II transmitted a position signal. Just over two minutes later — 193 seconds after the GRB began — the automated-response telescope at RIKEN in Japan took the first image of the afterglow. At 537, 721 and then 1,028 seconds, the 48-inch Palomar telescope in California took further images; telescopes at Kyoto and Bisai, Japan, photographed the fading afterglow around an hour after the burst. Follow-up by observatories all over the world produced an almost complete light curve (Fig. 2 on page 285), beginning at 193 seconds and ending several weeks later, when the afterglow finally disappeared.

This remarkable success brought its share of surprises. First, the early afterglow decayed much more slowly than had been predicted⁴, and more slowly than had been observed for GRB 990123. Second, the light curve showed a plateau (perhaps even a peak) around 0.1 days, and from 0.1 days onwards seemed to show significant wiggles superimposed on the power-law decay. This is unlike any other light curve seen before. With the exception of the early brightening of GRB 970508, all GRB afterglows have so far shown smooth, regular power-law decay. (In previous bursts the power-law decay was interrupted only by a 'jet break', corresponding to the onset of sideways propagation of the GRB.)

GRBs occur when an ultra-relativistic jet (a jet of particles moving at almost the speed of light) is slowed down and its enormous kinetic energy is dissipated. According to the popular internal–external shocks model (Fig. 1), an 'inner engine' — probably a collapsing massive star, or hypernova — generates the ultra-relativistic jet. Internal collisions, or shocks, within this jet produce the GRB itself, and the jet continues to flow outwards until it encounters the surrounding medium. Then, two external shocks are produced: the long-lasting forward shock propagates into the surrounding medium, exciting matter to produce first γ -rays and X-rays and later the optical and radio emission that constitutes most of the afterglow; a short-lived reverse shock propagates

back into the jet, producing optical and radio emission that decays rather quickly.

How does this model account for the unusual light curve recorded by Fox *et al.*?¹ In most cases, the external shocks would begin around one minute after the onset of the GRB while the internal shocks and the γ -ray emission are still going on. Hence, the very early afterglow and the reverse-shock emission overlap the late part of the GRB. The early part of the light curve may represent the emission from the reverse shock and the plateau a transition from the reverse to the forward shock⁵.

The late afterglow wiggles could be interpreted as the result of the shock wave encountering an external medium of variable density^{6–8}. But detailed calculations⁹ show that even an abrupt drop in density cannot explain the very steep decay of the afterglow seen around 0.3 days after the burst. Furthermore, such density variations are expected to have little influence on the X-ray band^{6,7}, and yet Fox *et al.*¹ find that fluctuations in the X-ray and optical light curves are correlated.

Alternatively, the shock wave's energy may have varied^{7,8}: an increase in shock-wave energy could explain the early slow decay, and an energy decrease would naturally produce a steep decline. A shock wave that has been slowed by the surrounding medium could be caught up by subsequent shocks, increasing the shock-wave energy¹⁰. If such 'refreshed' shocks do account for the early slower decay, then the actual energy release of GRB 021004 was significantly larger than implied by the observed γ -rays alone¹ — so if this burst is typical, then GRBs are even more energetic than we thought.

There is another possible explanation for energy modulation of the shock wave, which stems from the ultra-relativistic motion of the GRB jet^{7,11}. The emission from a radiating object moving at almost the speed of light is beamed into a very narrow cone along the line of motion. As the object slows down, the cone opens up to wider and wider angles. At first, during the GRB and the early afterglow, the shock wave is moving very fast and, because of this 'relativistic beaming', only a tiny fraction of the expanding jet is observed (only the part that is moving practically head-on towards the observer). As time passes and the blast wave slows down, a larger fraction of the jet is seen. The corresponding shock-wave energy is the average over the observed region and it may increase (or decrease) locally if the jet structure is inhomogeneous.

Both of these mechanisms are possible explanations for the energy increase seen by Fox *et al.*¹ in the early afterglow period. But the light curve for the later afterglow shows a period of steep decline, corresponding to energy decreases. As refreshed shocks can only increase the blast wave's energy, they cannot account for the later rapid decay.

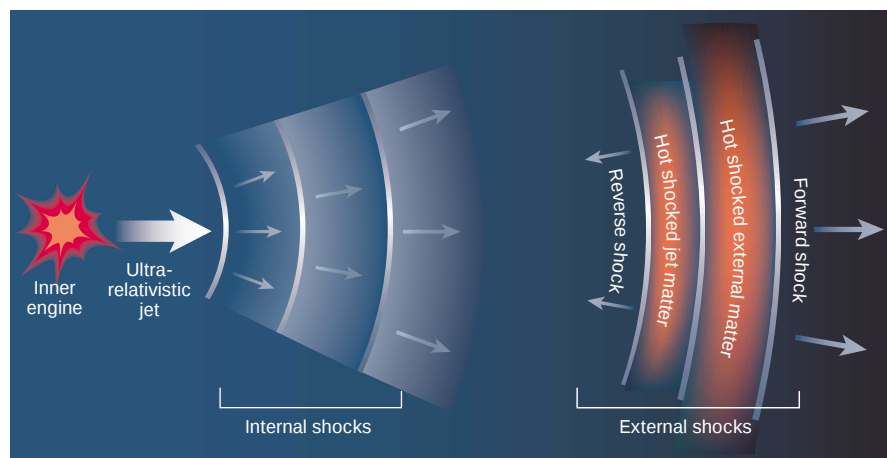


Figure 1 The internal-external shocks model. A gamma-ray burst (GRB) is thought to be driven by an 'inner engine', a cataclysmic event such as the collapse of a massive star. Inside an ultra-relativistic jet of particles thrown out from the explosion, internal shocks release a vast amount of energy in a burst of γ -rays. When the jet is slowed down by surrounding matter, external shocks are created: the forward shock that propagates further into space, and the reverse shock that is reflected back against the relativistic flow. Both types of shock waves heat the surrounding matter, producing an afterglow to the GRB.

It is more likely that a non-uniform jet structure is responsible for the observed pattern. Non-uniform structure in GRB jets has been long expected¹¹ — this may be the first evidence for its existence.

Tsvi Piran is at the Racah Institute for Physics, Hebrew University of Jerusalem, Jerusalem 91904, Israel.

e-mail: tsvi@phys.huji.ac.il

1. Fox, D. W. *et al.* *Nature* **422**, 284–286 (2003).

2. Akerlof, C. *Nature* **398**, 400–402 (1999).

3. <http://space.mit.edu/HETE/Welcome.html>

4. Sari, R. & Piran, T. *Astrophys. J.* **520**, 641–649 (1999).

5. Kobayashi, S. & Zhang, B. *Astrophys. J.* **582**, L75–L78 (2003).

6. Lazzati, D. *et al.* *Astron. Astrophys.* **396**, L5–L9 (2002).

7. Nakar, E., Piran, T. & Granot, J. *New Astron.* (in the press).

8. Heyl, J. & Perna, R. *Astrophys. J.* (in the press);

Preprint astro-ph/0211256 at <<http://arXiv.org>> (2002).

9. Nakar, E. & Piran, T. Preprint astro-ph/0303156 at

<<http://arXiv.org>> (2003).

10. Rees, M. J. & Mészáros, P. *Astrophys. J.* **496**, L1–L4 (1998).

11. Kumar, P. & Piran, T. *Astrophys. J.* **535**, 152–157 (2000).

Genomics

Gene expression meets genetics

Ariel Darvasi

Genetic analyses look for differences in gene sequence that could explain variation in physical traits. Gene-expression studies provide a snapshot of active genes. These approaches are now combined, to great effect.

There is an undeclared dispute among researchers who study complex traits — physical or behavioural characteristics of an organism that are dictated by combinations of more than one gene and the environment. On one side are classical geneticists, who look at the differences in DNA sequence between individuals and attempt to correlate them with physical and behavioural variations. On the other are proponents of gene-expression analysis, who focus on variations in the genes being switched on. On page 297 of this issue, Schadt and colleagues¹ describe how they brought these approaches together in a systematic, genome-wide analysis of the genetics of variation in gene expression. Their work provides support for a strategy by which complex traits could be studied at a greater scale and depth than is possible using either technique alone.

The idea of carrying out genome-wide genetic analyses of gene-expression data was introduced two years ago by Jansen and Nap², and Brem *et al.*³ were the first to apply this approach, in a study of budding yeast. The principles are outlined (for studying mice) in Fig. 1, overleaf. Crossing two progenitor strains, and subsequently crossing their offspring among themselves, produces a genetically variable population. Variations in DNA sequence across the genomes of this population are then analysed to identify their origin (that is, which one of the two progenitor strains). At the same time the population is studied to find out which genes are being expressed in different individuals, and to what degree. The expression level of each gene is then treated as a quantitative trait. Quantitative traits are determined by more than one gene and show a graded

variation across a population, such that the variation can only be measured quantitatively. Height and weight are typical examples, but it is perfectly reasonable to consider gene-expression levels as a quantitative trait, too.

Next, a statistical analysis is carried out to correlate DNA variations with gene-expression levels. A statistically significant correlation suggests that the gene (or genes) in the chromosomal region where the sequence variation occurs may account for some of the variation in gene expression. As the process is carried out for the entire genome, the results might highlight previously unknown gene–gene interactions, identify biochemical pathways and enable genetically like individuals to be grouped together. This last point in particular could be relevant to ‘personalized’ medicine—the development of drugs that are tailor-made for specific groups of patients.

Schadt *et al.*¹ have now carried out a genome-wide genetic analysis of gene expression — along the lines mentioned above — in three species, with an emphasis on mice. Initially, they produced a genetically varying population of 111 mice by crossing two commonly used inbred strains, C57BL/6J and DBA/2J. They then used a microarray chip representing 23,574 mouse genes to analyse the genes’ expression levels, and found that 7,861 are expressed differently in the two original strains. This provided the basis for measuring, as quantitative traits, the expression of each of the 7,861 genes in the population concerned. The authors also used a panel of more than 100 ‘microsatellite markers’ (sequence variations that differentiate the original strains) to analyse the DNA of the population.

Then, using standard statistical tools, Schadt *et al.* identified those genetic regions (loci) that can account for variation in the levels of gene expression. They term these loci eQTLs, for ‘expression quantitative trait loci’. Many of them were found to reside at the same chromosomal location as the gene whose expression level was being measured. As sequence variations in or near a gene can affect its expression, this observation was not entirely surprising, but rather serves as reassurance that the observed variation in gene expression is not random noise. More interestingly, Schadt *et al.* pinpointed eQTL ‘hotspots’ — chromosomal regions that contained significantly more eQTLs than would be expected by chance. This may point towards regulatory elements that affect the expression levels of several genes. The existence of such elements might explain, for example, how the expression levels of different genes in a given pathway are correlated. Another group of researchers has reached conceptually similar conclusions from a comprehensive eQTL analysis of a different mouse population⁴.

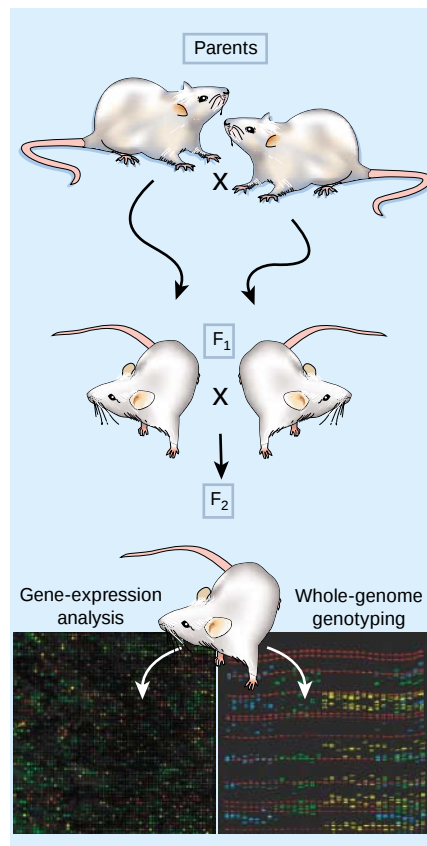


Figure 1 The genetics of gene expression. The approach shown here was first described by Jansen and Nap², and has now been extended to maize, mice and humans by Schadt *et al.*¹. Two distinct inbred strains of a species (the parental strains) are crossed to produce the F₁ generation, which in turn is crossed to produce the F₂ generation. Each F₂ individual has a random blend of chromosomal blocks originating from either parental strain. Using the F₂ population, any measurable set of physical or behavioural characteristics (phenotype) can be traced to a specific chromosomal location. This is achieved by correlating genetic (genotypic) and phenotypic variation in a population large enough to allow statistical significance. Schadt *et al.*¹ looked at the phenotype of variations in gene-expression levels across individuals in a population. An example of this type of approach is shown here. On the left, gene-expression analysis: each spot represents a different gene; different colours denote the expression level. On the right, genotyping: DNA fragments are coloured and run in an electrophoresis gel. Correlating the two allowed the authors to find chromosomal regions that affect gene expression.

An additional degree of sophistication can be introduced to such analyses by including a particular trait — a disease, for instance. Gene-expression data might help to define such a trait more accurately, generating genetically more homogeneous groups of individuals that have that characteristic. Schadt *et al.* looked at obesity, specifically the

mass of fat pads, in mice. They first divided mice into groups with high versus low fat-pad mass, and then analysed gene-expression data. They thereby subdivided the high-mass class into two subgroups. Genetic analysis of these subgroups then allowed the authors to identify QTLs (chromosomal regions that influence a quantitative trait, fat-pad mass in this case) that affect one subgroup but not the other.

The combination of gene-expression and genetic data could also have another use: it might help to identify candidate genes that affect a given trait. This could be achieved by looking for overlaps between differentially expressed genes and QTLs⁵, and by looking at eQTLs that are found in a common chromosomal region, as Schadt *et al.* suggest. Schadt *et al.* also broadened their study by looking at eQTLs in maize and humans, generalizing the approach to almost any organism in which both gene-expression profiling and genome-wide genetic analysis can be done efficiently.

There is little doubt that understanding the biology of complex organisms will require combinations of tools that examine more than one factor at a time^{6–8}, as exemplified by Schadt and colleagues’ work¹. But applying such tools presents challenges. For instance, including too many parameters carries the risk of producing false positives. And complicated experiments are more prone to misinterpretation. For instance, one of Schadt and colleagues’ most interesting findings is the eQTL hotspot phenomenon. But this might be somewhat falsely inflated because of the correlation of levels of gene expression among sets of genes, combined with the relatively high rate of false positives that is expected because of multiple testing (testing of several hypotheses) and a relaxed statistical threshold.

I expect that the combining of genetic information and gene expression will hasten the day when genomics delivers on its promise to improve health care. But we must continue striving to develop and apply sophisticated analytical tools for interpreting the vast, complex data sets that are being produced with modern genomic technologies.

Ariel Darvasi is at the Life Sciences Institute, Hebrew University of Jerusalem, Jerusalem 91904, and IDgene Pharmaceuticals Ltd, Jerusalem 91344, Israel.
e-mail: arield@cc.huji.ac.il

- Schadt, E. E. *et al.* *Nature* **422**, 297–302 (2003).
- Jansen, R. C. & Nap, J. P. H. *Trends Genet.* **17**, 388–391 (2001).
- Brem, R. B., Yvet, G., Clinton, R. & Kruglyak, L. *Science* **296**, 752–755 (2002).
- Manly, K. F. *et al.* 16th Int. Mouse Genome Conf. abstr. 88 (2002).
- Wayne, M. L. & McIntyre, L. M. *Proc. Natl Acad. Sci. USA* **99**, 14903–14906 (2002).
- Chong, L. & Ray, L. B. *Science* **295**, 1661 (2002).
- Kitano, H. *Science* **295**, 1662–1664 (2002).
- Jansen, R. C. *Nature Rev. Genet.* **4**, 145–151 (2003).

Optics

Ado about nothing much?

Markus Büttiker and Sean Washburn

Reports of signals travelling faster than the speed of light have captured the imagination of scientists and laymen alike. But there is an explanation for these effects that leaves Einstein's tenet intact.

In *Physical Review Letters*, Herbert Winful¹ presents a description of the "Nature of 'superluminal' barrier tunnelling", in answer to the controversy that has arisen from experiments^{2–5} in which pulses of light and microwaves appear to tunnel through a barrier at speeds faster than a reference pulse moves through a vacuum. This is an apparent violation of Einstein's founding assumption of relativity: nothing can move faster than light through a vacuum. Can light go faster than light? What does 'faster' really mean in this context? As each pulse is in fact a 'packet' of waves, the central issue is the definition of the speed of the wave packet.

Tunnelling occurs when a wave impinges on a thin barrier of opaque material and some small amount of the wave 'leaks' through to the other side. The superluminal experiments that caused controversy were performed with a lattice of layers of transparent and opaque materials arranged so that waves of some frequencies are reflected (through destructive interference) but other frequencies pass through the lattices. All materials behave like this to some extent: the glass in a window is transparent to optical frequencies (so you can keep an eye on children playing in the yard) but relatively opaque to infrared frequencies (so your greenhouse can trap these waves and warm your plants in the winter sunshine).

The lattice of opaque and transparent layers forms one arm of the experiment. A pair of pulses is separated: one is sent through the lattice on a tunnelling path, and the other

is routed through open space along a reference path. Whether the pulses arrive simultaneously at the end of their paths is checked through their interference. Initially the path length of the reference arm is set so that the photons arrive at the same instant when the lattice is not in place. But when the lattice is installed in the tunnelling arm, the length of the reference arm has to be shortened for the photons to arrive simultaneously — implying that the pulse has traversed the tunnelling arm at a speed greater than that of the pulse travelling unimpeded through empty space.

But in fact this is not profoundly troubling, as is clear from the following argument⁶. According to quantum mechanics, light can be thought of as a wave or as a stream of particles (photons). The lattice can be represented as regions of empty (transparent) space interleaved with identical, thin opaque layers at regular separations. The barriers are so thin that the light-pulse photons spend virtually no time in them. Incident photons are either transmitted through a particular barrier or reflected back. Without question, the fastest route possible is the one in which the photon passes through all of the barriers without any reflection; the second fastest route is if the photon is reflected backwards at one opaque layer then reflected forwards again with no further reflections; and so on. As the tunnelling is a result of the destructive interference in the lattice that makes the whole lattice appear opaque, the leading photon, which has not been reflected even once, is attenuated the least. Photons with a

single reflection are attenuated more but still less than photons that have suffered two or more reflections.

In all cases, the pulse that emerges from the tunnelling process is greatly attenuated, and 'front-loaded' — only the leading edge of the incident pulse survives the tunnelling event without being severely attenuated to the point that it cannot be detected. If we measure the speed by the peak of the pulse, it looks faster than the incident pulse (Fig. 1). But that is just an artefact of our definition of speed as marked by the arrival of the peak in the pulse.

As there is no sharp beginning to a pulse, we cannot declare — as we could for an athlete crossing a finish line — the instant of its arrival at a certain point. The question is fundamentally meaningless. We can only watch the rising edge of the pulse and try to recognize what is arriving. So Winful¹ rightly concludes that "the rather prompt appearance of the peak at the output should not be cause for concern [as] the incident peak did not propagate to the output". He further concludes that this notion of the tunnelling time is meaningless, an opinion shared with a number of scientists^{7,8} (including one of ourselves⁹).

Einstein and his contemporaries knew that pulses in complex media are not bound by the laws of relativity. The pulse is a group of waves, interfering with one another constructively (or, as in the discussion above, not quite completely destructively) that moves at the 'group velocity' of the medium. This group velocity is largely disconnected from the speed of the waves that compose it, just as a cork floating in the sea does not progress forward as fast as the swells on which it bobs. Sommerfeld pointed out⁷ that the key question is how long it takes to recognize what is arriving at the output. If a signal pulse is sent at some instant, how much time elapses before the observer at the output can identify the signal being sent? This time lapse is called the signal propagation time, and its speed, the signal speed. The signal speed is always subluminal — even when the apparent pulse speed or group velocity is superluminal. Recognition of the signal might involve reading the frequency¹⁰ (which requires several wavelengths to pass the observer), and the observer must disentangle the pulse shape, which in turn makes the detection slow and complicated. The upshot is that signal recognition admits only subluminal effective speeds for all imagined pulse shapes and recognition schemes.

For the reasons described above and some more subtle ones related to the necessarily gradual process of switching on or sending the pulse, the time delay associated with the peak of the pulse must be replaced by more sophisticated time-keeping methods^{9,10}. One cannot label the rising and falling edges of the pulse with (for example)

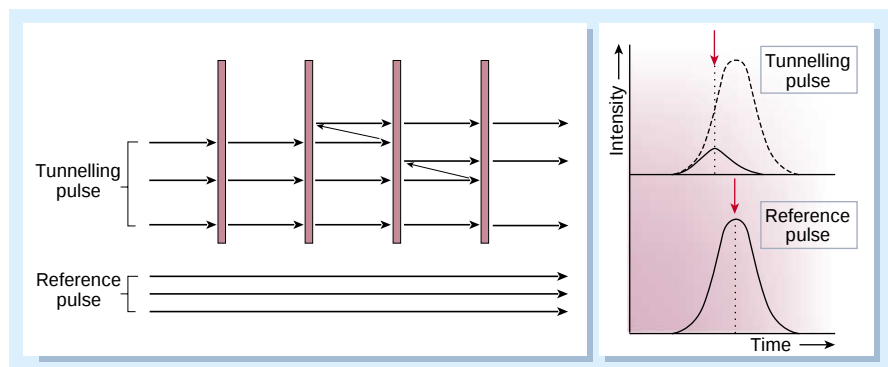


Figure 1 Faster than the speed of light? Experiments have compared a light pulse tunnelling through a lattice of opaque and transparent material with a reference pulse travelling, unimpeded, in a vacuum, and the results suggest that the tunnelling pulse travels faster — faster than the speed of light. But in fact this is just an artefact of the experiment. Photons in the tunnelling pulse are randomly transmitted or reflected by the lattice layers, and the final pulse that emerges is attenuated. If the time of arrival of the pulses is judged from the peak of the pulse, the tunnelling pulse appears to have travelled faster.

different polarizations without destroying the interference^{11,12}. Perhaps the best solution is to use the waves that make up the pulse itself as the clock to measure the delay⁹. By modulating (for example) the size of the barrier, one can distinguish clearly between pulses that pass through much faster than the time over which the barrier oscillates and those that pass through so slowly that the barrier oscillates many times before they escape. By changing the frequency of the barrier modulation one can then time the tunnelling process.

Alternatively, the amount of rotation of spins or dipoles could be used as a clock, where the rotation is caused by the spin¹³ or dipole¹⁴ tunnelling through a region of magnetic field (the Larmor effect) or electric field (the Faraday effect). As such a clock has nothing to do with the shape of the rising edge of the wave packet, it might give a more meaningful measure of the tunnelling time. Experiments of this kind would provide a key to the question of whether superluminal tunnelling is an important development, or just a misnomer.

On the other hand, in a practical sense, the issue is moot. As a scientist has pointed out, with reference to the commercialization of the tunnelling devices as fast links in optical and microwave transmission, it does not matter what we think: people are selling them already. The presumption in such commercial ventures is that even short-range, inefficient, repetitive transmission may help to provide increased speed in some manipulations of pulses or particles in future computing or communication. From the point of view of physics, however, it seems that the class of such applications will be narrow at best.

Markus Büttiker is in the Department of Theoretical Physics, University of Geneva, 24 quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland.
e-mail: buttiker@karystos.unige.ch

Sean Washburn is in the Department of Physics and Astronomy, Applied and Materials Sciences, University of North Carolina, Chapel Hill, North Carolina 27599-3255, USA.
e-mail: sean@physics.unc.edu

1. Winful, H. G. *Phys. Rev. Lett.* **90**, 023901 (2003).
2. Steinberg, A. M., Kwiat, P. G. & Chiao, R. Y. *Phys. Rev. Lett.* **71**, 708–711 (1993).
3. Nimtz, G., Enders, A. & Spieker, H. J. *Phys. I (France)* **4**, 565–570 (1994).
4. Spielmann, C., Szpöcs, R., Stingl, A. & Krausz, F. *Phys. Rev. Lett.* **73**, 2308–2311 (1994).
5. Ranfagni, A., Mugnai, D., Fabeni, P. & Pazzi, G. *Appl. Phys. Lett.* **58**, 774–776 (1991).
6. Japha, Y. & Kurizki, G. *Phys. Rev. A* **53**, 586–590 (1996).
7. Sommerfeld, A. *Phys. Z.* **8**, 841–842 (1907).
8. Brillouin, L. *Ann. Phys.* **44**, 203–240 (1914).
9. Büttiker, M. & Landauer, R. *Phys. Rev. Lett.* **49**, 1739–1742 (1982).
10. Büttiker, M. & Thomas, H. *Ann. Phys. (Leipzig)* **7**, 602–617 (1998).
11. Deutch, J. M. & Low, F. E. *Ann. Phys.* **288**, 184–202 (1993).
12. Steinberg, A. M., Kwiat, P. G. & Chiao, R. Y. *Fund. Phys. Lett.* **7**, 223–239 (1994).
13. Büttiker, M. *Phys. Rev. B* **27**, 6178–6188 (1983).
14. Gasparian, V., Ortuño, M., Ruiz, J. & Cuevas, E. *Phys. Rev. Lett.* **75**, 2312–2315 (1995).

Developmental biology

A hairy situation

Yann Barrandon

A key step in hair-follicle development is the rearrangement of epithelial stem cells. It seems that reduced production of an adhesion protein, through the concerted action of two signalling pathways, is crucial for this process.

It's often difficult to tell when two people are siblings — despite their common origins and upbringing, they are usually very different. Similarly, it is not always obvious when distinct cell types have been produced from the same precursor cells. For instance, would you have guessed that mammary glands, sweat glands and hair follicles all develop from the same discrete accumulation of stem cells resting in the primitive epidermis, the outermost cell layer of an embryo? The first step in the production of different cell types from these stem cells is the formation of an epithelial bud, a small cluster of cells that bulges out from the epidermis (Fig. 1). On page 317 of this issue, Jamora *et al.*¹ describe how they used hair-follicle development as a model system to show that decreased production of the adhesion factor E-cadherin — known for its role in mediating cell–cell contacts — is key for epithelial-bud development. Moreover, this downregulation of E-cadherin requires the sequential action of two signals, Wnt and noggin, that were previously thought to be unlinked.

During development, the formation of an epithelial bud is the first evidence that some stem cells of the primitive epidermis have been instructed to form a hair follicle, sweat gland or mammary gland. To reach this new spatial arrangement, epithelial cells must first change the way in which they interact with their neighbours, and break their attachments to the underlying matrix. Only then can they move about to take up a new rearrangement and overall tissue shape. Cells achieve this by adjusting their repertoire of cell–matrix and cell–cell adhesion molecules.

The molecules that instruct epidermal stem cells to form an epithelial bud have long been sought². For hair follicles, overwhelming evidence points to a role for proteins in the cell nucleus that regulate gene expression (transcription factors), especially the Lef1 transcription complexes^{2–4}. Indeed, forced expression of Lef1 in mice results in the formation of extra hair follicles, and these can even develop from tissues that are normally hairless. By contrast, loss of the Lef1 gene results in severely impaired mammary gland, tooth and hair development. So what might activate Lef1 in the stem cells and, equally importantly, how does Lef1 then promote follicle formation? We know that pathways triggered by the Wnt protein can switch on Lef1 complexes in

other contexts, through stabilization of the Lef1-activating protein β -catenin, but is this relevant for follicle development? Jamora *et al.*¹ show that it is. They find that, in response to Wnt, Lef1 transcription complexes downregulate the expression of the E-cadherin gene in epithelial buds, and so provide a direct link between Lef1 and the cell's adhesion machinery. Furthermore, the authors challenge the view that Lef1 complexes always act as transcriptional activators in response to Wnt signalling³; they find that Wnt proteins, via β -catenin, instead cause Lef1 to act as a transcriptional repressor.

A further twist to this tale is that Wnt signalling alone is not sufficient for Lef1 activation. So what might be the second signal that acts with Wnt? There are several candidates: hair-follicle development also requires the coordinated expression of members of various protein families — the fibroblast growth factor, bone morphogenetic protein (BMP), Sonic hedgehog and Notch families, as well

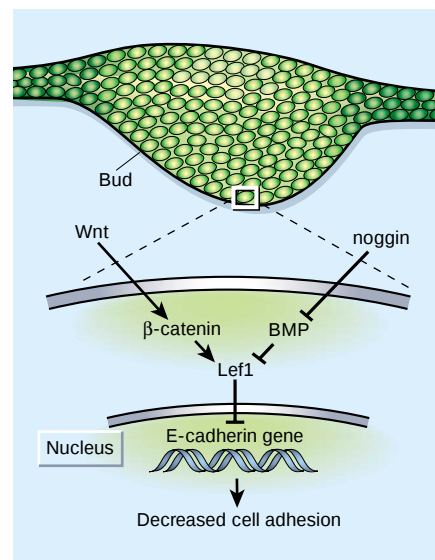


Figure 1 Making a bud. The formation of an epithelial bud from the epidermal layer of an embryo (top part of figure) is an early step in the development of mammary glands, hair follicles and sweat glands. Jamora *et al.*¹ have found that decreased expression of the adhesion protein E-cadherin is needed for bud formation. The darker green indicates high E-cadherin expression (epidermis); the lighter green indicates low E-cadherin expression (bud). The lower part of the figure shows the signalling pathways that Jamora *et al.* have found to lead to decreased E-cadherin expression.



SEAWIFS PROJECT; NASA/GODDARD SFC; ORBIMAGE

Meteorology

Getting the wind up

Once research into global change took off on a large scale, few phenomena related to climate escaped measurement. Researchers have dragged unwieldy equipment up mountains in the tropics to drill into the glaciers found there. And the deep ocean has been probed by fleets of sophisticated floats. But wind speeds in the eyewall of hurricanes, where the storm is at its most intense, have long defied observation. On page 279 of this issue, however, Mark Powell and his colleagues present just such

measurements (*Nature* **422**, 279–283; 2003).

The data — 331 vertical wind profiles from 15 tropical cyclones — were obtained by dropwind sondes from the Global Positioning System. The sondes are launched from aircraft at altitudes of 1.5–3 km; they relay their position and measure pressure, temperature and humidity as they fall through the atmosphere and drift with the wind. Tropical cyclones — such as Hurricane Alberto, which occurred in August 2000 and is shown in the satellite

image here — derive much of their energy from sea–air exchange. For forecasting purposes it is essential to have accurate information about relevant parameters such as the exchange of momentum, which partly depends on the drag factor exerted by the state of the sea surface.

One of Powell and colleagues' findings is that maximum wind speeds occur at an altitude of about 500 m. More notably, however, they find that one assumption used in predicting the intensity and

consequences of hurricanes is incorrect. Previously, in the absence of observations for wind speeds above 25 m s^{-1} , levels of increasing drag with increasing wind speed were extrapolated to high wind speeds. But now it seems that above hurricane force — about 33 m s^{-1} — a layer of foam and bubbles from breaking waves develops that reduces drag and effectively lets the hurricane glide over the sea. In consequence, air–sea exchange in hurricanes will need to be reassessed. **Heike Langenberg**

as several transcription factors². Moreover, hair-follicle development is severely impaired when the *noggin* gene, which encodes a BMP inhibitor, is deleted⁵. Jamora and colleagues now link BMP and Wnt signalling by showing that activation of the Wnt pathway results in the formation of β -catenin-activated Lef1 transcription complexes — and consequently in decreased E-cadherin expression — only when combined with the inhibition of BMP signalling by *noggin*¹.

So is the downregulation of E-cadherin indeed necessary for hair-follicle formation? Cadherins are calcium-dependent proteins that mediate cell–cell adhesion, often as part of complexes called adherens junctions, and they are known to be involved in development and tissue maintenance⁶. In the skin, E-cadherin is expressed in the developing epidermis (whereas P-cadherin becomes predominant later in the growing hair follicle)⁷. Previous work has shown that the loss of some intracellular cadherin-binding proteins severely impairs hair-follicle formation^{4,8}, but no one had looked at whether E-cadherin itself is required in this process. Jamora *et al.* now test this and show that the forced expression of E-cadherin in the developing skin results in increased adhesiveness and abnormal epithelial buds. So, the loss of E-cadherin (and, by implication, changes in adherens

junctions) appears to be crucial for epithelial buds to begin forming. This supports a model in which downregulation of the E-cadherin gene in stem cells that receive *noggin* and Wnt signals facilitates cell movement, rearrangement and bud formation. It is also consistent with previous observations that the loss of E-cadherin expression or function can lead to enhanced invasive behaviour of other cells^{6,9}.

Together with previous work¹⁰, the results of Jamora and colleagues provide strong evidence that dynamic changes in the composition of adherens junctions are important for the development of skin appendages such as hair follicles. The results may also be relevant to the development of other tissues and organs from epithelial buds (for example, teeth, lungs and limbs). That said, many questions remain unanswered. For instance, the pelage follicles (guard hairs) that occur in mice develop in a Lef1-independent manner. Yet the new results imply that downregulation of E-cadherin expression is common to all developing hair follicles. This hints that other transcription regulators can also repress E-cadherin expression in epithelial cells during follicle formation (and we know of good candidates, such as members of the Snail family^{11,12}). In addition, epithelial buds do not develop randomly, but follow a precise geometrical pattern. What governs this?

Another important question is what determines whether an epithelial bud contributes to a hair follicle, sweat gland or mammary gland. Here there are likely to be species-specific mechanisms, because we know, for example, that hair follicles and sweat glands form in mutually exclusive regions in mice, whereas in humans both cell types can develop from neighbouring stem cells. But are these fates specified early in bud formation or later during bud expansion? The choices an epithelial bud makes are clearly intricate and await investigation. ■

Yann Barrandon is at the School of Life Sciences, Swiss Federal Institute of Technology (EPFL), and the Department of Experimental Surgery, Vaud University Hospital (CHUV), 1011 Lausanne, Switzerland.
e-mail: yann.barrandon@epfl.ch

- Jamora, C., DasGupta, R., Kocieniewski, P. & Fuchs, E. *Nature* **422**, 317–322 (2003).
- Fuchs, E. & Raghavan, S. *Nature Rev. Genet.* **3**, 199–209 (2002).
- Eastman, Q. & Grosschedl, R. *Curr. Opin. Cell Biol.* **11**, 233–240 (1999).
- Huelsken, J. *et al.* *Cell* **105**, 533–545 (2001).
- Botchkarev, V. A. *et al.* *Nature Cell Biol.* **1**, 158–164 (1999).
- Takeichi, M. *Science* **251**, 1451–1455 (1991).
- Hirai, Y. *et al.* *Development* **105**, 271–277 (1989).
- Vasioukhin, V. *et al.* *Cell* **104**, 605–617 (2001).
- Gumbiner, B. M. *J. Cell Biol.* **148**, 399–403 (2000).
- Perez-Moreno, M., Jamora, C. & Fuchs, E. *Cell* **112**, 535–548 (2003).
- Cano, A. *et al.* *Nature Cell Biol.* **2**, 76–83 (2000).
- Batlle, E. *et al.* *Nature Cell Biol.* **2**, 84–89 (2000).

Obituary

Jean Brossel (1918–2003)

Jean Brossel passed away on 4 February 2003 in Périgueux, Dordogne, the city of his birth. He was one of the founding fathers of the modern French school of quantum optics and co-founder, with Alfred Kastler, of the Laboratoire Kastler Brossel at the Ecole Normale Supérieure, in Paris.

Brossel entered the Ecole Normale in 1938, completing his studies in 1945 (after a two-year interruption due to the Second World War). Physics research in France was devastated by the war, and lagged behind the scientific developments in Britain and the United States. To catch up, Kastler advised Brossel to begin his research in the United Kingdom. Brossel took his advice and joined Samuel Tolansky's group in Manchester, in the laboratory headed by Patrick Blackett. There he learned how to apply interferometry to the study of surfaces, and to use Fabry–Perot interference techniques for the measurement of hyperfine structure in electronic transitions inside the atom. Throughout his life, Brossel acknowledged the value of the expertise he acquired in Tolansky's laboratory.

On Brossel's return from England, Kastler recommended that he go to the United States to work for a PhD with Francis Bitter at the Massachusetts Institute of Technology. Bitter had just informed Kastler about exciting new ideas for Doppler-free spectroscopy (in which Doppler broadening of optical emission lines does not limit the resolution) that might extend radio-frequency studies of atomic resonances to excited states. Bitter's idea turned out to be incorrect, but it proved a fruitful starting point for Brossel that later led to his invention of the method of 'double resonance': in this Doppler-free method, the atoms are subjected to simultaneous irradiation with light and a radio-frequency field, and resonant effects are detected by measuring the polarization of the fluorescent light emitted by the atoms (equivalent to the angular momentum of the photons), instead of its frequency (the energy of the photon).

While Brossel was finishing his thesis with Bitter, Kastler followed a similar line of work in Paris and devised 'optical pumping', by which atoms in their lowest-energy (or ground) state acquire angular momentum through successive cycles of absorption of polarized optical radiation, followed by spontaneous emission. The final result is a selective

population of magnetic sublevels inside the ground state.

In 1951, Kastler and Brossel founded a research group together at the Ecole Normale — what became the Laboratoire Kastler Brossel (LKB). Their primary objective was to demonstrate optical pumping experimentally and to explore the possibilities it created. Excellent students from the Ecole Normale soon joined the group (Bernard Cagnac, Jacques Winter, Jacques Blamont and Jean-Pierre Barrat, among many others). This field of research proved to be a treasure trove of new physics. Optical pumping was demonstrated in 1952, and was followed by the observation of multiphoton resonances, where several photons, instead of a single one, are absorbed by the atoms at the same time — Brossel provided a physical interpretation of the selection rules for the polarization of the radio frequency in terms of photon angular momentum. Then, in work overseen by

Physicist who opened up new perspectives in quantum optics

Brossel and Kastler, came coherent radiation trapping and polarization transfer, the general distinction between diagonal and off-diagonal elements of the density matrix (populations and coherences) and the discovery of light shifts. After Kastler received the Nobel prize in 1966 for his work on optical methods for studying resonances of the atom, he often expressed his deep regret that Brossel had not shared it with him.

Brossel encouraged young scientists to pursue new directions of research in his laboratory. This research included work on atomic masers and magnetometers, parity-violation in atomic physics, cavity electrodynamics and Doppler-free two-photon spectroscopy. Perhaps the best-known application of optical pumping is in atomic clocks, and the work done at LKB, in particular on light shifts, was crucial to their development. Improvements in the optical pumping of rubidium atoms led to the creation in 1969 of a magnetometer that was sensitive to magnetic fields at the level of 10^{-9} gauss. In 1997, Brossel's colleague and former student Claude Cohen-Tannoudji shared the Nobel prize for his work on ultra-cold atoms and laser cooling.

Although Brossel was interested in the applications of his research, he was



sceptical about application-driven research in universities. He argued that technical breakthroughs are the fruit of fundamental research driven by pure scientific curiosity, and rarely of application-oriented programmes. A good example is atomic clocks, which sprang from fundamental studies in atomic physics rather than from a drive to improve on mechanical or quartz clocks. Another example is optical pumping in helium-3. First performed 20 years ago by King Walters and Laird Schearer in the United States, optical pumping in this isotope was refined at LKB, motivated by the study of quantum effects in optically pumped gases. Now optical pumping in helium-3 forms the basis of a tool for the medical imaging of lung cavities in asthma patients.

Jean Brossel was for many years the head of the physics department at the Ecole Normale, to which he attracted brilliant and active colleagues. Teaching was always important to him, in particular as a way of attracting the best students to physics. But even at the peak of his administrative responsibilities, he never abandoned 'hands-on' work in the laboratory. For instance, he continued to prepare glass cells containing atomic gases, sometimes with elaborate coatings to prevent wall relaxation — an art in which no one could match his skill. He will be remembered by his colleagues and former students with great admiration.

Franck Laloë

Franck Laloë is in the Laboratoire Kastler Brossel, Département de Physique de l'ENS, 24 rue Lhomond, 75005 Paris, France. e-mail: lalo@lkb.ens.fr

LABORATOIRE KASTLER BROSSEL

Cosmology

The helium legacy of dead stars

Mon. Not. R. Astron. Soc. (in the press); Preprint astro-ph/0302285 at <<http://arXiv.org>> (2003)

Findings from studies of the first stars in the Universe may upset standard models of the Big Bang. According to R. Salvaterra and A. Ferrara, these objects, called Population III stars, might have produced an appreciable fraction of the helium-4 (^4He) found in interstellar space and thought previously to have been created mainly in the Big Bang itself.

Already, models of element formation in the Big Bang don't quite match up with observations of the amounts of these 'primordial' elements in the Universe: there seems to be rather less ^4He and ^7Li than expected. If much of this ^4He was made by fusion of hydrogen in Population III stars, that puts the predictions even further astray.

When subatomic particles first condensed from the Big Bang, they formed mostly ^1H . About 10% of this matter was ^4He , with smaller amounts of ^2H , ^3He and ^7Li . But ^4He would also have been made from ^1H in the first stars. The trouble is, these stars are now dead, so their abundance must be estimated indirectly. Salvaterra and Ferrara estimate it from the near-infrared background (NIRB) radiation, the glow emitted by cosmic gas warmed by starlight ever since the first stars shone. The NIRB has been measured recently, and it implies that Population III stars were more productive than previously thought.

Philip Ball

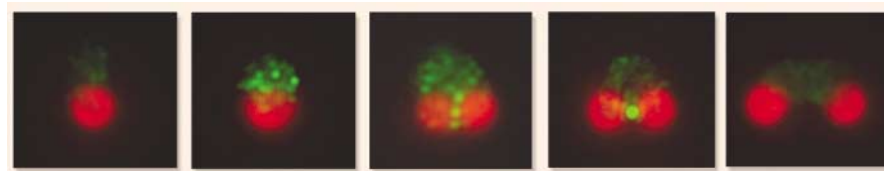
Plant cell biology

Ancient rings

Plant Cell 15, 655–665 (2003)

Mitochondria are energy-generating organelles found in all organisms; chloroplasts are the site of photosynthesis in plants. Both are thought to have originated long ago when simple, bacterium-like cells were engulfed by more complex cells. During cell proliferation these organelles must be replicated and separated, and a protein with bacterial origins, FtsZ, is involved in chloroplast division. Mitochondria in higher organisms lack FtsZ and instead use the protein dynamin, implying that chloroplasts kept the ancient division system while mitochondria gained a new one.

But Shin-ya Miyagishima *et al.* now report that in the primitive red alga *Cyanidioschyzon merolae*, chloroplasts use dynamin as well. One type of dynamin, CmDnm1, was already known to be involved in mitochondrial division in *C. merolae*. Through genome searches, Miyagishima *et al.* uncovered a related



Dynamin dynamics: a chloroplast (red) is severed by a ring of dynamin protein (green).

dynamin, CmDnm2, which during cell division migrates to the chloroplast to form a ring that constricts around the chloroplast's 'waist' (as pictured here).

A relative of FtsZ had already been found to participate in mitochondrial division in *C. merolae*. So in this primitive plant, dynamin and FtsZ help both chloroplasts and mitochondria divide. This hints that the ancestors of these organelles likewise used rings of both proteins.

Christopher Surridge

Ecology

Nuclear reptiles

Mol. Ecol. 12, 743–752 (2003)

Lizards can be faithful too — D. O'Connor and R. Shine have discovered that the Australian black rock skink, *Egernia saxatilis*, forms long-term pairs to raise its offspring. The authors found from genetic tests that more than 80% of skink groups include one female adult and one male adult, with at least some male–female pairs and offspring remaining together throughout the three-year duration of the research project. Families share a territory, but spend most of their time apart.

Egernia is the most sociable lizard known — 15 of the 29 species in this genus live in some form of family group. But these tend to be larger and more diffuse than those of *E. saxatilis*, which is the first reptile found to have a 'nuclear family' structure of a pair and offspring living together for more than one year. Fidelity in this species is not perfect, however, in that a few groups contained more than two adults, and a fifth of clutches had offspring fathered by an external male.

John Whitfield

Surface chemistry

Riddle of the stones

J. Am. Chem. Soc. 125, 2854–2855 (2003)

Kidney stones, which cause excruciating pain when they lodge in the urinary tract, are a mystery. They are typically formed by the crystallization and build-up of calcium oxalate on the surfaces of the kidney, but it isn't clear how the crystals become attached to renal cells. Xiaoxia Sheng and colleagues have gone fishing for clues at the molecular scale, using an atomic force microscope (AFM).

There is evidence that carboxylate groups in biological macromolecules

might be the hooks that snag the growing crystals to cells of the renal epithelium. To test this notion, the researchers coated the tip of an AFM with long-chain organic molecules carrying a carboxylate (or other organic groups for comparison) at the free end, and waved it over the surface of single calcium-oxalate crystals.

They found that the adhesion force is several times larger for carboxylates than for other end groups, such as methyl or alcohols. And adhesion is disrupted when poly(aspartic acid) is added to the solution, suggesting that this polymer sticks to the crystal surface and exhausts all the binding sites. That squares with the fact that carboxylate-rich macromolecules — similar to poly(aspartic acid) — in urine help suppress crystal attachment to renal cells.

Philip Ball

Developmental biology

Tools for single-cell analysis

Dev. Cell 4, 383–393 (2003)

During development, immature cells are often committed to becoming specific mature cell types long before any physical signs of this commitment can be seen. Changes in gene-expression patterns are presumably the key, but most research has analysed populations of cells and so has been unable to identify any differences between individual cells.

Ming-Ko Chiang and Douglas A. Melton have now used a modified form of the polymerase chain reaction, together with DNA microarrays, to study the genes expressed by single pancreas cells from mouse embryos. The microarrays are custom-made to light up when active genes from a cell form a match with any of the known pancreas genes on the microarray chip. Chiang and Melton identified six subtypes of developing pancreas cells, each expressing a unique combination of genes, and in some subtypes they discovered combinations of four genes — *Pds1* to *Pds4* — that were not previously known to be expressed in the pancreas.

The researchers propose a model of pancreas development in which the early pancreas contains a heterogeneous mix of cells, with genetically distinct cell types becoming apparent over time. They also suggest that this type of analysis could help to characterize cell subtypes in other biological systems.

Helen R. Pilcher

Radiographic imaging with cosmic-ray muons

Natural background particles could be exploited to detect concealed nuclear materials.

Despite its enormous success, X-ray radiography¹ has its limitations: an inability to penetrate dense objects, the need for multiple projections to resolve three-dimensional structure, and health risks from radiation. Here we show that natural background muons, which are generated by cosmic rays and are highly penetrating, can be used for radiographic imaging of medium-to-large, dense objects, without these limitations and with a reasonably short exposure time. This inexpensive and harmless technique may offer a useful alternative for detecting dense materials — for example, a block of uranium concealed inside a truck full of sheep.

In X-ray radiography, the intensity of an image pixel is determined by the attenuation of the incident beam caused by absorption and scattering — the maximum mean free path for photons is about 25 g cm^{-2} for all materials, corresponding to less than 2 cm of lead. For thicker objects, it is better to use a different type of radiography that is based on the interaction of charged particles with matter by multiple Coulomb scattering. The many small interactions add up to yield an angular deviation that roughly follows a gaussian distribution,

$$\frac{dN}{d\theta_x} = \frac{1}{\sqrt{2\pi}\theta_0} e^{-\frac{\theta_x^2}{2\theta_0^2}}$$

with the width, θ_0 , related to the scattering material through its radiation length, L_0 , as follows:

$$\theta_0 = \frac{13.6}{\beta c p} \sqrt{\frac{L}{L_0}} [1 + 0.038 \ln(L/L_0)]$$

where p is the particle's momentum in MeV c^{-1} and βc is its velocity². The radiation length decreases rapidly as the atomic number of a material increases, and θ_0 increases accordingly: in a layer 10 cm thick, a 3-GeV muon will scatter with an angle of 2.3 milliradians in water, 11 milliradians in iron and 20 milliradians in lead. By tracking the scattering angles of individual particles, the scattering material can be mapped.

Our new technique relies on the scattering of atmospheric muons produced by primary cosmic rays. Muons are the most numerous cosmic-ray particles at sea level, moving at a rate of about $10,000 \text{ m}^{-2} \text{ min}^{-1}$ in horizontal detectors³. These particles are highly penetrating: a typical cosmic-ray

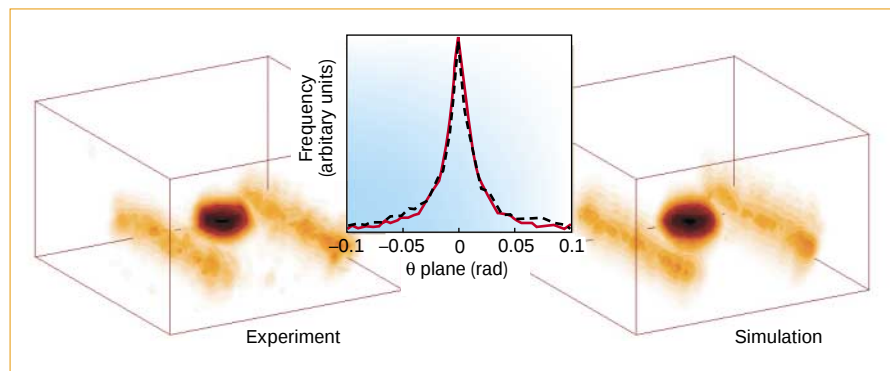


Figure 1 Radiographic imaging with muons of a test object (left) and the reconstructed image of its Monte Carlo simulation (right). The test object is a tungsten cylinder (radius, 5.5 cm; height, 5.7 cm) on a plastic ($35 \times 60 \times 1 \text{ cm}^3$) plate with two steel support rails. The tungsten cylinder and the iron in the rails are clearly visible in both the experiment and simulation reconstructions. Inset, the widths of the scattering distributions for tracks passing through the tungsten target are very similar for the experimental and simulated data.

muon of energy 3 GeV will penetrate more than $1,000 \text{ g cm}^{-2}$ (10 m of water, for example).

To demonstrate the concept of muon radiography, we developed a small-scale experimental system with four drift chamber detectors⁴ spaced 27 cm apart. Each detector has an active area of $60 \times 60 \text{ cm}^2$ and records particle tracks at two positions in each of two orthogonal coordinates. The upper pair of detectors records the tracks of incident muons, and the lower pair records the scattered tracks. A tungsten cylinder was used as a test object, supported by a plastic plate and steel support beams. The tungsten is clearly visible in the reconstructed image, and the steel support beams are also evident (Fig. 1, left).

We also developed a Monte Carlo simulation code that generated cosmic-ray muons and propagated them through a test volume. The reconstructed images are indistinguishable from those obtained experimentally (Fig. 1), and the scatter angles of the simulated muons from the different materials (tungsten, lexan and steel) are consistent with the measured angles.

Simulation of larger, more complex objects demonstrates that we can reliably detect a $10 \times 10 \times 10 \text{ cm}^3$ uranium object inside a large metal container full of sheep in 1 min of exposure. We conclude that cosmic-ray muons show promise as an inexpensive, harmless probe for radiography of medium-to-large objects, such as commercial trucks, passenger cars or sea containers. Our experimental results and simulations demonstrate the ability to reconstruct complex objects and to detect dense material of high atomic number hidden in a much larger volume of material of low atomic number, using only the natural

flux of muons. This method is suitable for a range of practical applications in which radiography of dense objects with low radiation dose is required — for example, in surveillance for cross-border transport of nuclear materials.

Konstantin N. Borozdin, Gary E. Hogan, Christopher Morris, William C. Priedhorsky, Alexander Saunders, Larry J. Schultz, Margaret E. Teasdale
Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
e-mail: kbor@lanl.gov

1. Roentgen, W. C. *Nature* **53**, 274 (1896).
2. Hagiwara, K. et al. *Phys. Rev. D* **66**, 01001 (2002).
3. Grieder, P.K.F. *Cosmic Rays at Earth* (Elsevier Science, Amsterdam, 2001).
4. Atencio, L. G., Amann, J. F., Boudrie, R. L. & Morris, C. L. *Nucl. Instr. Methods Phys. Res.* **187**, 381–386 (1981).

Competing financial interests: declared none.

Palaeo-oceanography

Deepwater variability in the Holocene epoch

The conversion of surface water to deep water in the North Atlantic results in the release of heat from the ocean to the atmosphere, which may have amplified millennial-scale climate variability during glacial times¹ and could even have contributed to the past 11,700 years of relatively mild climate (known as the Holocene epoch)^{2–4}. Here we investigate changes in the carbon-isotope composition of benthic foraminifera throughout the Holocene and find that deep-water production varied on a centennial–millennial timescale. These variations may be linked to surface and atmospheric events that hint at a contribution to climate change over this period.

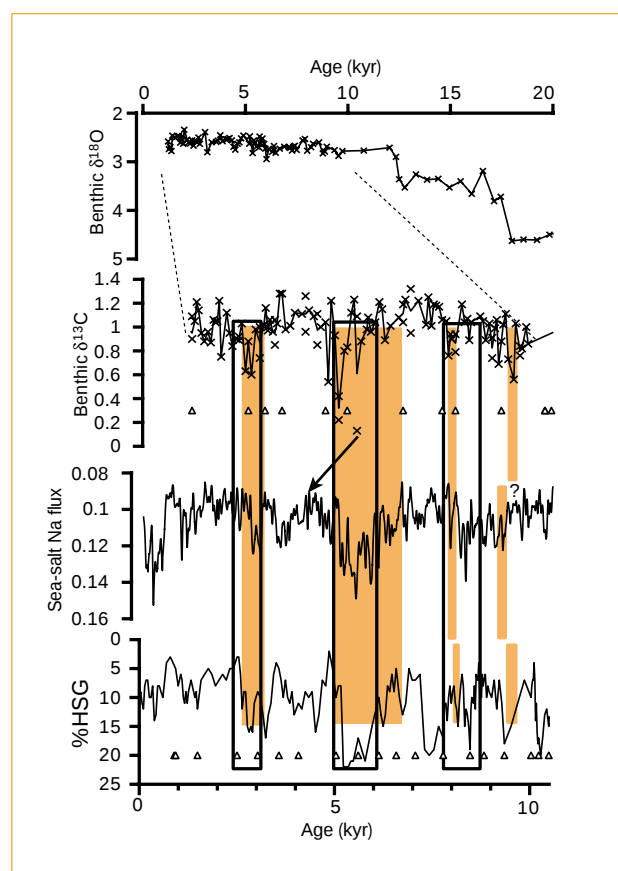


Figure 1 Holocene climate records, top to bottom: benthic $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ (North Atlantic deepwater contribution) records from site 980 (ref. 12, and see supplementary information); GISP2 sea-salt sodium flux⁸ (increasing, calm to windy), and percentage of haematite-stained grains (%HSG) in core V29-191 (ref. 2). Triangles, dated levels in sediment cores; arrow marks the onset and intensification of the 5-kyr event. Accelerator mass-spectrometer radiocarbon dates converted to calendar age¹³ from site 980 provided the chronology for the Holocene (see supplementary information). Accumulation rates average about 25 cm kyr^{-1} in the Holocene, roughly double that in nearby core V29-191. The average interval between samples is about 100 yr from 9.7 to 1.2 kyr. Shading shows low- $\delta^{13}\text{C}$ events (see text) and possible correlative events in the other records; a more tenuous correlation is denoted by the question mark. Rectangles, extreme winter-like conditions deduced from statistical analysis of the full set of palaeochemical indicators from GISP2 (ref. 8).

We used an established deepwater proxy parameter — the carbon-isotope composition ($\delta^{13}\text{C}$) of the epifaunal benthic foraminifera *Cibicidoides wuellerstorfi*^{5,6} — to evaluate Holocene deepwater variability. Variations in the $\delta^{13}\text{C}$ of the total amount of CO_2 in bottom waters, which is accurately recorded by *C. wuellerstorfi*, can be used to monitor variations in the contribution of high- $\delta^{13}\text{C}$ North Atlantic Deep Water (NADW) relative to low- $\delta^{13}\text{C}$ Southern Ocean Water (SOW) to a site. We worked on sediment from Ocean Drilling Project site 980 (55° N , 15° W ; depth, 2,179 m), on the Feni Drift in the subpolar northeastern Atlantic.

Holocene $\delta^{13}\text{C}$ values (Fig. 1, and see supplementary information) show marked millennial oscillations around the modern value of about 1‰ (ref. 7). Values range from a high of 1.3‰ to values lower than 0.6‰, indicating times of enhanced and reduced NADW contribution, respectively. The largest reductions in the relative NADW contribution occurred around 9,300 years ago (9.3 kyr), and 8.0, 5.0 and 2.8 kyr ago. Smaller events occurred more frequently.

The most pronounced feature of the Holocene is a trend of decreasing relative NADW contribution that began at about 6.5 kyr and culminated with a minimum at around 5 kyr. Statistical analysis of the full set of chemical records from the Greenland Ice Sheet Project 2 (GISP2) ice core indicates that meteorological conditions from

6.1 to 5.0 kyr were especially winter-like (characterized by an expanded polar vortex) at high latitudes; this is consistent with high values of sea-salt sodium (Fig. 1) during this millennial-scale event⁸.

A high relative abundance of haematite-stained grains in a nearby core indicates a large proportion of cold, fresh, ice-bearing surface water from north of Iceland². All three records indicate that the 5-kyr event was the most severe climate event of our Holocene study interval. Apparently contemporary variations in benthic $\delta^{13}\text{C}$ and sea-salt sodium levels indicate that linked, century-scale variability may have occurred during this event.

The GISP2 palaeochemical records indicate enhanced winter-like conditions from 2.4 to 3.1 kyr, and from 8.8 to 7.8 kyr, although to a lesser extent than in the 5-kyr event⁸. The more recent of these cold intervals corresponds precisely to the reduced NADW contribution at around 2.8 kyr. The 8-kyr $\delta^{13}\text{C}$ minimum occurs late in the 8.8–7.8-kyr cold interval, but corresponds to a brief maximum in sea-salt sodium. The age of the 9.3-kyr event is poorly constrained, but it may correspond to one of several younger or older maxima in sea-salt sodium flux. Three of the four low- $\delta^{13}\text{C}$ events correspond to maxima in the percentage of haematite-stained grains (increased incursion of rock-bearing ice from north of Iceland), suggesting a possible surface–deepwater linkage.

Prior geochemical evidence for deep-water reduction in the Holocene was limited to events outside our study interval. Low $\delta^{13}\text{C}$ values during the Little Ice Age⁴, the most recent of the Holocene millennial cold events⁸, and during an early Holocene event at ~ 10.3 kyr (ref. 9) hint at a linkage between millennial climate and deep water. Downcore variations in sedimentological indices also indicate this possibility^{10,11}.

The new North Atlantic benthic $\delta^{13}\text{C}$ record unambiguously demonstrates that NADW varied on centennial–millennial time scales during the Holocene. The most significant Holocene event of reduced NADW contribution, which occurred at 5 kyr, was correlated to winter-like atmospheric conditions at high northern latitudes⁸, and to incursion of sea ice from north of Iceland². Similarly, a more recent cold event (at 3.1–2.4 kyr) was associated with a reduced NADW contribution.

Evidence for a climate–deepwater linkage during the earlier events is weaker, and may indicate an increasing sensitivity of deepwater to surface forcing from the Early to the Late Holocene. Further well-dated deepwater proxy records are needed to test this possibility.

Our study raises other important issues that can be addressed through data collection and modelling. The amplitude of the largest Holocene $\delta^{13}\text{C}$ fluctuation near 5 kyr, for example, is similar to those of $\delta^{13}\text{C}$ oscillations during earlier glacial and deglacial events¹². This suggests that significant variations in the relative NADW contribution can occur in the absence of forcing by large ice sheets, and that NADW may be more sensitive to surface forcing than was previously imagined.

Delia W. Oppo*, Jerry F. McManus*, James L. Cullen†

*Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02540, USA

e-mail: doppo@whoi.edu

†Department of Geological Sciences, Salem State College, Salem, Massachusetts 01970, USA

1. Broecker, W. S., Bond, G. & Klas, M. *Paleoceanography* **5**, 469–477 (1990).
2. Bond, G. C. et al. *Science* **294**, 2130–2136 (2001).
3. Alley, R. B. et al. *Geology* **25**, 483–486 (1997).
4. Keigwin, L. D. & Boyle, E. A. *Proc. Natl Acad. Sci. USA* **97**, 1343–1346 (2000).
5. Curry, W. B., Duplessy, J. C., Labeyrie, L. D. & Shackleton, N. J. *Paleoceanography* **3**, 317–341 (1988).
6. Duplessy, J. C. et al. *Paleoceanography* **3**, 343–360 (1988).
7. Kroopnick, P. M. *Deep-Sea Res.* **32**, 57–84 (1985).
8. O'Brien, S. R. et al. *Science* **270**, 1962–1964 (1995).
9. Bond, G. C. et al. *Science* **278**, 1257–1266 (1997).
10. Bianchi, G. G. & McCave, I. N. *Nature* **397**, 515–517 (1999).
11. Chapman, M. R. & Shackleton, N. J. *Holocene* **10**, 287–291 (2000).
12. McManus, J. F., Oppo, D. W. & Cullen, J. L. *Science* **283**, 971–975 (1999).
13. Stuiver, M. & Braziunas, T. F. *Radiocarbon* **35**, 137–189 (1993).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Reduced drag coefficient for high wind speeds in tropical cyclones

Mark D. Powell*, Peter J. Vickery† & Timothy A. Reinhold‡

* National Oceanic and Atmospheric Administration, Atlantic Oceanographic and Meteorological Laboratory, Hurricane Research Division, Miami, Florida 33149, USA

† University of Western Ontario, Boundary Layer Wind Tunnel, London, Ontario N6A 5B9, Canada

‡ Clemson University, Department of Civil Engineering, Clemson, South Carolina 29634, USA

The transfer of momentum between the atmosphere and the ocean is described in terms of the variation of wind speed with height and a drag coefficient that increases with sea surface roughness and wind speed. But direct measurements have only been available for weak winds; momentum transfer under extreme wind conditions has therefore been extrapolated from these field measurements. Global Positioning System sondes have been used since 1997 to measure the profiles of the strong winds in the marine boundary layer associated with tropical cyclones. Here we present an analysis of these data, which show a logarithmic increase in mean wind speed with height in the lowest 200 m, maximum wind speed at 500 m and a gradual weakening up to a height of 3 km. By determining surface stress, roughness length and neutral stability drag coefficient, we find that surface momentum flux levels off as the wind speeds increase above hurricane force. This behaviour is contrary to surface flux parameterizations that are currently used in a variety of modelling applications, including hurricane risk assessment and prediction of storm motion, intensity, waves and storm surges.

In strong winds, momentum exchange at the sea surface is described by a sea-state-dependent drag coefficient (C_d). The behaviour of C_d in tropical cyclones has never been observed, so it is currently based on extrapolations from field measurements in much weaker wind regimes. These extrapolations describe an increase in C_d with wind speed and are currently employed in applications such as: forecasting the intensity and track of tropical cyclones in numerical weather prediction models¹; diagnosing surface wind speeds from reconnaissance flight-level observations^{2,3}; determining the geographic distribution of extreme wind loads and loss with probabilistic models⁴ for building design and insurance ratemaking, respectively; and specification of the wind field forcing for storm surge^{5,6} and wave forecasts⁷.

Under neutral stability in strong winds, the vertical variation of mean wind speed is controlled by the roughness of the sea surface. Relevant surface layer quantities are sea surface roughness length (Z_0), friction velocity (U_*), and the neutral stability 10-m wind speed (U_{10}) and drag coefficient (C_d). Assuming a neutrally stable surface layer, the mean wind speed (U) increases logarithmically⁸ with height (Z).

$$U = (U_*/k) \ln(Z/Z_0) \quad (1)$$

where k is a constant of about 0.4, and U_* is the friction velocity related to the surface momentum flux (τ)

$$\tau = \rho U_*^2 = \rho C_d U_{10}^2 \quad (2)$$

and ρ is the air density. The roughness length described by Charnock⁹ is:

$$Z_0 = \alpha U_*^2/g \quad (3)$$

where α is a constant of range 0.015–0.035. Estimates of U_* , Z_0 and C_d are computed by measurement of the momentum flux using eddy correlation and inertial dissipation methods or by using the mean wind speed profile with equations (1) and (2).

Field studies based primarily on eddy correlation and inertial dissipation measurements have examined C_d and Z_0 dependence on U_{10} and sea state in winds below hurricane force^{10–15}. Indirect estimates of C_d for hurricanes^{16–18} have been made from angular

momentum and vorticity budgets, but errors are large. C_d depends on Z_0 , U_{10} , wave age (ratio of local friction velocity to phase speed of dominant spectral component) and wave steepness^{19–22}. Influences on Z_0 include steepness of wind waves, surface current velocity, and relative directions of wind, waves, current and swell. These studies suggest that swell and wind waves from different directions and fetches modulate the local Z_0 independently of U_{10} . Individual waves may distort the overlying wind field²³ by setting up speed-up and separation zones over and to leeward of the wave crest, respectively. Estimates of the vertical extent of this effect range from a few centimetres²⁴ to three wave heights²⁵.

Recent measurements of directional wave spectra²⁶ gathered by NOAA research aircraft in Hurricane Bonnie of 1998 depict the classic conceptual model²⁷ with the longest and largest (shortest and smallest) waves in the right front (left rear) quadrant, but show a complex pattern with multiple wave modes propagating at large and sometimes conflicting angles to the local wind, with a tendency for steeper waves in the right rear quadrant.

Wind measurement by GPS sonde

High-resolution wind profile measurements are now possible owing to the development of the Global Positioning System dropwind-sonde (GPS sonde)²⁸. From 1997–1999, 331 wind profiles were measured in the vicinity of the hurricane eyewalls in the Atlantic, and Eastern and Central Pacific basins (Table 1). The eyewall contains the strongest winds of a tropical cyclone and is identified as the ring of convection typically surrounding the relatively clear ‘eye’ depicted in satellite and radar imagery.

The GPS sonde is launched from reconnaissance or research aircraft at altitudes of 1.5–3 km or higher and falls at a vertical speed of 10–15 m s^{−1}, while sampling pressure, temperature, humidity and position every 0.5 s. A sonde launched in the eyewall takes several minutes to reach the surface while drifting (relative to the storm centre) tangentially 10–15 km and radially hundreds of metres. Wind speeds calculated from the motion of the GPS sonde have an accuracy of 0.5–2.0 m s^{−1}, and height is typically estimated to within 2 m. GPS sonde measurements were smoothed by a 5-s low-pass filter to remove fluctuations associated with undersampled scales and noise due to satellite switching. Extreme

turbulence and intense rainfall contribute to signal interruptions or failures of GPS sondes to report values all the way to splash at the sea surface. Limited comparisons of GPS sondes to NOAA moored buoy and Coastal-Marine Automated Network platform observations²⁹ suggest GPS sonde winds are within 3.5 m s^{-1} of nearby buoy measurements at moderate wind speeds.

Wind profiles were examined in a composite sense (as a normalized profile containing all measurements gathered in the lowest 3 km), and also as a function of mean boundary layer (MBL) wind speed. For the MBL analysis, individual profiles were organized into five groups according to the MBL wind speed, defined as the mean wind speed of all profile observations below 500 m. The five groups corresponded to MBL wind speeds in the ranges of $30\text{--}39 \text{ m s}^{-1}$ (72 profiles), $40\text{--}49 \text{ m s}^{-1}$ (105 profiles), $50\text{--}59 \text{ m s}^{-1}$ (55 profiles), $60\text{--}69 \text{ m s}^{-1}$ (61 profiles) and $70\text{--}85 \text{ m s}^{-1}$ (38 profiles). Near-surface (8–14 m) winds sampled by the GPS sondes ranged from 21 to 67 m s^{-1} . Each group was ordered vertically into height bins chosen to provide the highest resolution where the wind shear is the greatest. Variability associated with mesoscale, convective, and undersampled turbulent scales was removed by averaging all profiles in a given wind speed group.

Wind profile scaling

Scaled winds and heights were considered for constructing a normalized wind profile. A scaling height quantity was not selected because of the lack of a consistent objective method for estimating boundary layer height (thus the arbitrary definition of MBL) and variability or uncertainty of other possible quantities, including depth of the inflow layer, gradient wind level and height of the maximum wind speed. Planetary boundary layer (PBL) scaling heights determined from potential temperature or specific humidity mixed layer depths could differ by a factor of two.

The ‘gradient’ wind level is frequently used in engineering applications to represent the height at which the ‘free atmosphere’ flows above the boundary layer. Ideally, the gradient wind flows parallel to lines of constant pressure (isobars) and represents the source of momentum that can be transported to the surface in gusts. A ‘gradient height’ is difficult to assign in a tropical cyclone, because ‘gradient balance’³⁰ may apply to winds as high as 3 km above the surface. A gradient height based on zero inflow (assuming inflow represents frictional surface effects of the PBL) yields values 1–3 km above thermodynamic measures of boundary layer height.

The MBL was selected to normalize the wind measurements because it usually contains the maximum wind speed measured in the profile, is not greatly affected by multiple maxima and minima, and also contains all the boundary layer measurements. The MBL

wind was within 10% of the maximum mean wind in all groupings, so it was consistent with the concept of a ‘gradient’ wind as a source of momentum for gusts affecting the surface.

Mean wind profile

The normalized composite wind profile in Fig. 1 contains all (over 126,000) individual 0.5-s samples from 331 profiles comprising all MBL groups. The lower 200 m of the profile shows a logarithmic increase of mean wind speed with height, reaching a maximum near 500 m. Wind speeds decreased above 500 m owing to the weakening of the horizontal pressure gradient with height in the warm core of the hurricane^{31,32}. Increased variability of the individual wind measurements above 600 m contradicts the concept of a ‘free atmosphere’ flow above the PBL.

Most of this variability is probably caused by convective scale features in the eyewall as well as the location of the GPS sonde launch relative to the flight-level wind maximum. The outward tilt of angular momentum surfaces with height³³ (eyewall tilt) places the location of the maximum wind at typical launch altitudes (3 km) as much as 1–3 km radially outward (relative to the storm centre) from the location of the maximum wind at the surface. GPS sondes launched radially outward from the 3-km wind maximum would fall into weaker MBL winds, whereas those launched radially inward from the 3-km maximum would tend to fall into stronger MBL winds.

Estimation of surface winds and gusts

The relationship of the surface and typical flight-level winds to the MBL can be used to estimate surface winds based on reconnaissance flight-level wind measurements. U_{10} is about 78% (standard deviation 8%) of the MBL. The 1.5-km flight-level wind is about equal (13%) to the MBL for MBL groups $<60 \text{ m s}^{-1}$, and then decreases to 96% (12%) and 93% (10%) for the 60–69 and 70–85 m s^{-1} MBL groups, respectively. The ratio of the 3-km flight-level wind to the MBL decreases from 110% (37%), to 97% (20%), to 90% (14%), to 87% (12%) and to 84% (11%) as the MBL wind speed group

Table 1 GPS sonde wind profiles in tropical cyclones

Storm	Year	Profiles	Tropical storm	Cat1	Cat2	Cat3	Cat4	Cat5
Bonnie	1998	84	–	0	84	0	0	0
Bret	1999	11	–	0	1	3	7	0
Danielle	1998	4	–	4	0	0	0	0
Dennis	1999	32	2	8	22	0	0	0
Dora	1998	8	–	4	3	1	0	0
Erika	1997	22	–	0	0	22	0	0
Eugene	1999	1	–	1	0	0	0	0
Floyd	1999	58	1	4	10	20	23	0
Georges	1998	22	–	0	0	9	13	0
Gert	1999	7	–	0	1	0	6	0
Guillermo	1997	6	–	0	0	6	0	0
Irene	1999	2	–	0	2	0	0	0
Jose	1999	1	–	0	1	0	0	0
Lenny	1999	22	–	5	3	6	8	0
Mitch	1998	51	–	0	3	5	37	6
Totals		331	3	26	130	72	94	6

Number of individual GPS sonde wind profiles in tropical cyclones by storm and by Saffir–Simpson damage potential scale category (Cat) as determined by the National Hurricane Center ‘best track’.

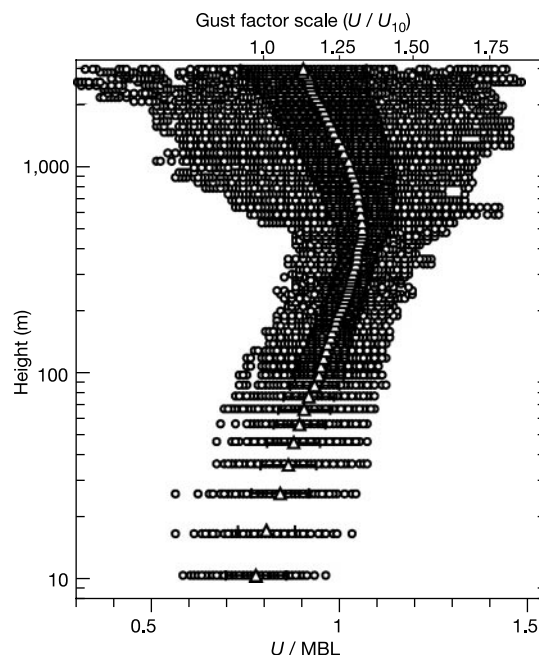


Figure 1 GPS sonde wind speed measurements normalized by mean boundary layer wind. Individual measurements (open circles), mean (open triangles) and standard deviation (horizontal bars) are shown for each height bin. An alternative scale of U/U_{10} is also shown, to represent the surface gust factor.

increases successively from 30–39 to 70–85 m s^{-1} .

It must be stressed that these factors are only relevant for deep-water, open-ocean conditions. A variety of field experiments conducted in shallow water with shoaling waves suggest that higher roughness (and smaller surface wind reduction factors) would be experienced adjacent to coastal areas. This is of special concern now that GPS sonde information is influencing a reanalysis of the major US hurricane landfalls (for example, <http://www.noaanews.noaa.gov/stories/s966.htm>). Such reassessments are premature until sufficient numbers of GPS sonde profiles become available to distinguish between open-ocean conditions and shoaling and breaking wave effects in coastal areas.

The ratio of the surface wind to MBL was used to create an 'alternative' gust factor scale (ratio of the individual wind speed measurements to the mean surface wind speed) in Fig. 1. For mean winds above 25 m s^{-1} observed by coastal anemometers during hurricane landfall³⁴, gust factors (ratio of the peak 3-s gust to the 10-min mean) average at 1.5 with the largest gust factors at about 1.8. For marine exposure winds over 25 m s^{-1} , the mean gust factor³⁵ is around 1.4. Convective gusts^{36,37} are believed to represent the transport and acceleration of extreme winds to the surface through downdrafts with gust factors of over 2.0. Observations of convective gusts in tropical cyclones are extremely rare because power outages and anemometer failures make it difficult to preserve continuous wind speed records. The highest wind observations at 10 m in Fig. 1 suggest gust factors less than 1.3, which is in agreement with values measured by NOAA buoy platforms in hurricanes. Examination of tropical cyclone anemometer records typically shows peak gust values of the order of three standard deviations in excess of the mean. The envelope of minima and maxima in Fig. 1 may, however, not be comparable to lulls and gusts measured by conventional anemometers, because filtering during post-processing effectively smoothes out high-frequency fluctuations.

Drag coefficient and surface roughness

Analysis of eyewall thermodynamic soundings³⁸ document well-mixed layers of potential temperature and specific humidity consistent with well-developed boundary layers. Mean wind profiles for each MBL group (Fig. 2) were logarithmic in the lowest 200 m and then levelled off slightly with a peak near 500 m. The log profile is consistent with a neutral stability surface layer in which the vertical distribution of momentum is controlled by surface roughness, as described by equation (1). The profiles suggest a decrease in the elevation of the maximum mean wind as the MBL wind speed increases. This behaviour is consistent with a shallower boundary layer in stronger storms.

The lowest 100–150 m of each MBL group in Fig. 2 was fitted by a least-squares line to determine the intercept (on a natural log height scale) and slope as a measure of Z_0 and k/U_* respectively, according to equation (1). The drag coefficient was computed from equation (2). Values of U_* , Z_0 and C_d for four estimates of surface layer depth are shown in Fig. 3. The 10–150-m layer tended to have the smallest error bars. All fits explained over 98% of the variance.

The group in the range 70–85 m s^{-1} contained too few low-level samples to determine reliable estimates of slope and intercept. Surface layer parameter dependence on wind speed (Fig. 3) showed that U_* increased with U_{10} up to 40 m s^{-1} before levelling off, and that Z_0 and C_d initially increased as surface winds approached hurricane force (33 m s^{-1}). For $U_{10} < 40 \text{ m s}^{-1}$, U_* and Z_0 behaviour (Fig. 3a, b) was very similar to that described by Large and Pond¹³, although Z_0 decreased with $U_{10} > 40 \text{ m s}^{-1}$ and the Z_0 values became much less than those predicted by the Charnock⁹ and Large and Pond relationships. The magnitudes of C_d (Fig. 3c) for $U_{10} < 40 \text{ m s}^{-1}$ were consistent with numerous investigations³⁹, several budget studies conducted in tropical cyclones, and recent flume and annulus experiments (M. A. Donelan, personal com-

munication; see also ref. 40), although the tendency was for a decrease in C_d for $U_{10} > 33 \text{ m s}^{-1}$. The most remarkable result was the large decrease in Z_0 and C_d as U_{10} increased to 51 m s^{-1} , which was not indicated by any of the investigations shown in Fig. 3c. However, no observations previously existed for such wind speeds (the values shown from previous investigations are extrapolations).

A possible explanation for the transition in Z_0 and C_d as U_{10} increased from 40 to 51 m s^{-1} is the development of a sea foam layer⁴¹ at the air–sea interface. Surface winds above hurricane force (34 m s^{-1}) create streaks of bubbles on the sea surface combined with patches of foam 20–50 m wide caused by steep wave faces breaking and being sheared off by the wind (Fig. 4a). As the wind approaches 50 m s^{-1} , the sea becomes completely covered by a layer of foam and it is difficult to discern individual wave-breaking elements in the reduced visibility from spray and rain (Fig. 4b). Experiments with a bubble annulus⁴² suggest that bubble layers in salt water impede the transfer of momentum from the wind. As winds increase above 40 m s^{-1} , it is speculated that increased foam coverage could progressively form a 'slip' surface at the air–sea interface. In addition to the possible effects of foam, sea spray is

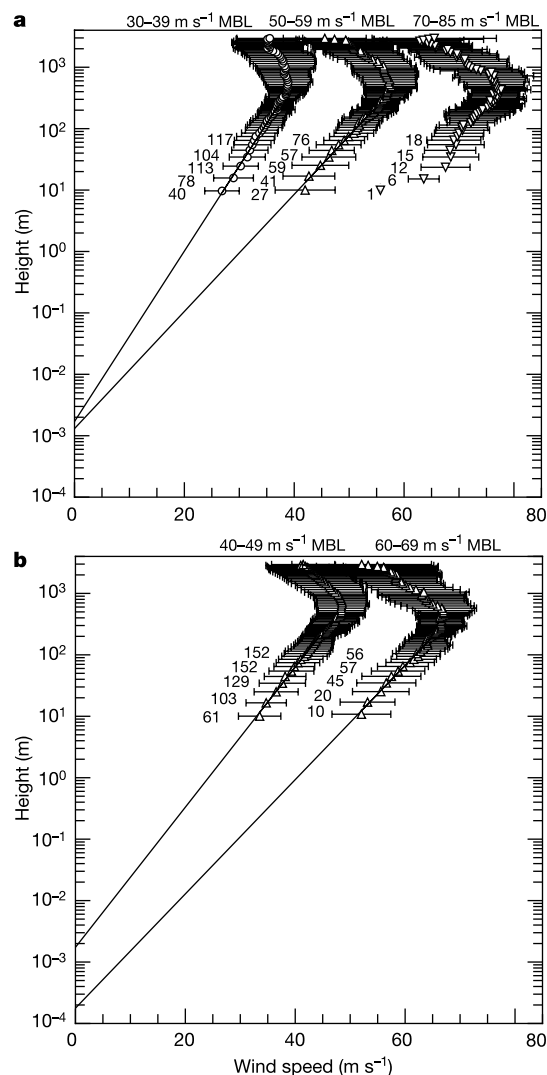


Figure 2 Mean wind profiles by MBL group. Symbols and horizontal bars represent mean and standard deviation for each vertical bin, numbers adjacent to each level represent the number of samples within each of the lowest five height bins. **a**, 30–39 m s^{-1} (72 profiles), 50–59 m s^{-1} (55 profiles) and 70–85 m s^{-1} (38 profiles) MBL groups. **b**, 40–49 m s^{-1} (105 profiles) and 60–69 m s^{-1} (61 profiles) MBL groups.

hypothesized to significantly influence the transfer of momentum, heat and moisture in tropical cyclones⁴³.

Remote sensing measurements from the NOAA WP-3D aircraft⁴⁴ provide independent evidence supporting the C_d and Z_0 behaviour observed in the GPS sonde profiles. Scatterometer measurements of

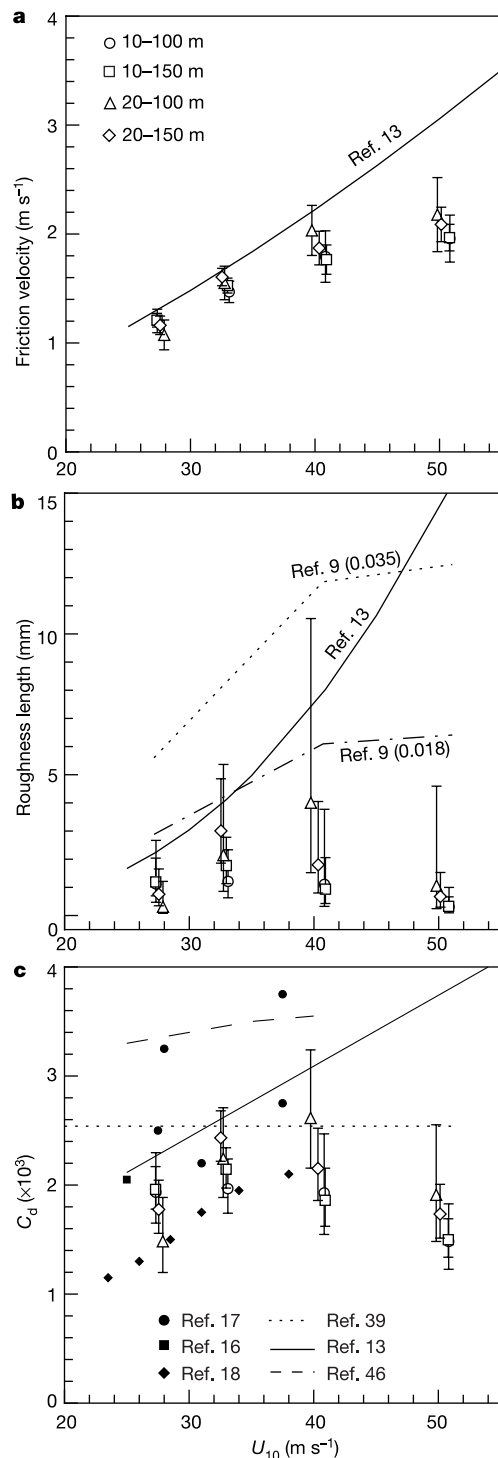


Figure 3 Surface momentum exchange quantities as a function of U_{10} . Vertical bars represent the range of estimates based on 95% confidence limits and symbols (see key in **a**) show estimates based on the depth of the surface layer used for the fit. **a**, U_* , from GPS sondes (symbols) and Large and Pond relationship¹³ (solid line). **b**, Z_0 from GPS sonde profiles (symbols) with Large and Pond¹³ (solid line) and Charnock⁹ relationships for comparison (dotted line, $\alpha = 0.035$; dashed-dot line, $\alpha = 0.018$). **c**, C_d , with relationships and values from tropical cyclone budget studies^{13,16–18,39,46}.

normalized radar cross-section (σ_0) due to Bragg scattering from surface capillary waves ‘saturate’ or fail to continue to increase at threshold wind speeds over 40 m s⁻¹. A levelling off of σ_0 with increasing U_{10} might be associated with the spacing and steepness of wind waves such that the wind ‘sees’ an effective surface that represents the tops of the separation zones between the waves (M. A. Donelan, personal communication). A reduction in C_d with increasing U_{10} is also supported by estimates of the angular momentum budget and the implied frictional dissipation in Hurricane Frederic⁴⁵ but the wind speeds were much less than those considered here. A field investigation of ocean response to Hurricane Gilbert⁴⁶ determined C_d from the near-inertial energy flux across the top of the oceanic thermocline. Although the C_d values exceeded those described here, a levelling-off behaviour is indicated.

An alternative scenario to describe the Z_0 and C_d behaviour implied by the mean profiles concerns speculation on the existence of linear coherent features in the form of helical rolls⁴⁷. At low levels, GPS sondes could be preferentially drawn into a convergent and downward part of a helical roll circulation that would contain the strongest winds, contributing to a profile with smaller shear (and lower Z_0 and U_*) near the surface. Validation of such features will require simultaneous multiple Doppler radar measurements of the same sampling volume.

A final possible explanation for the shape of the mean wind profiles and resulting surface layer parameters involves sampling strategy. Most GPS sondes are launched strategically to place them near the surface wind maximum. They are usually dropped radially inward from the flight-level wind maximum. The GPS sonde trajectories are generally tangential until they reach the lowest 300–500 m, at which point the radial (inward relative to the storm centre) component of the wind increases. Hence, the GPS sonde

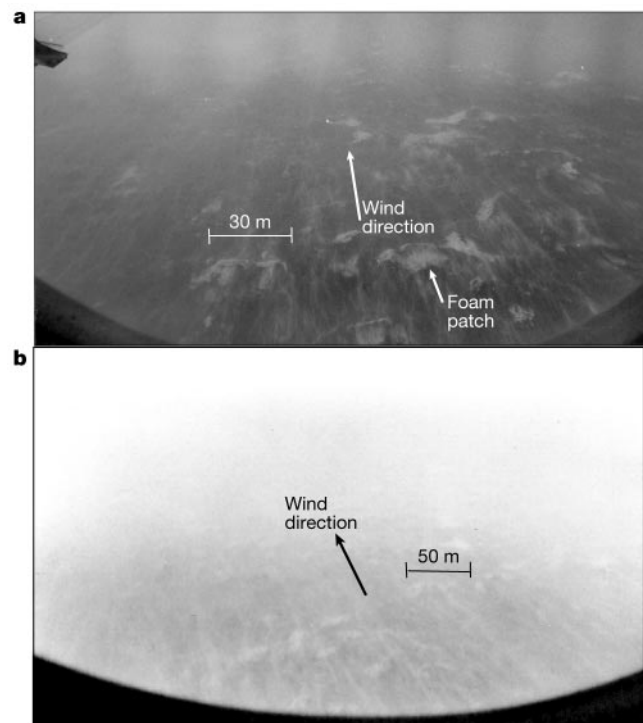


Figure 4 Sea state photographs during an eyewall penetration in Hurricane Ella on 1 September 1978. **a**, At 15:21 UTC (universal time coordinated) from 253-m altitude with flight-level wind speed of 46 m s⁻¹ (starboard wing is at upper left). **b**, at 15:23 UTC from 450-m altitude with flight-level wind speed of 55 m s⁻¹. Arrows show the wind direction and scales suggest approximate sizes of wave features. Photographs taken by M.D.P. aboard the NOAA WP-3D research aircraft.

may be moving horizontally towards the surface wind maximum, causing the reported winds to increase while falling closer to the surface (thus contributing to weaker wind shear). However, it seems just as likely that a GPS sonde could fall through the location of the surface wind maximum well before reaching the surface, and then be moving towards the relatively weak eye.

Until now, there have been no observations of wind profiles, surface stress, Z_0 and C_d in winds much above 25 m s^{-1} . Our findings provide compelling evidence that logarithmic mean wind profiles exist in tropical cyclones and that the surface stress, Z_0 and C_d are much less than previously thought in winds above 40 m s^{-1} . Heat and moisture transfer coefficients have a U^* dependence and may also be reduced. Such winds comprise less than 1% of a mature tropical cyclone's circulation, but have a crucial role in enhanced surface fluxes that supply enthalpy to the eyewall. The amount of enthalpy in the eyewall is directly related to the intensity of the tropical cyclone⁴⁸.

Surface layer parameterizations used in tropical cyclone motion, intensity, wave and storm surge prediction and risk assessment may not be valid in high wind, fetch-limited conditions over the open ocean. In the light of possible effects on heat and moisture transfer, shoaling effects in shallow water, and azimuth-dependent sea state and environmental wind-shear-induced asymmetries in the wind field, we consider that continued collection and examination of GPS sonde profiles is warranted. □

Received 18 September 2002; accepted 4 February 2003; doi:10.1038/nature01481.

- Bender, M. A., Ginis, I. & Kurihara, Y. Numerical simulations of tropical cyclone-ocean interaction with a high-resolution coupled model. *J. Geophys. Res.* **98**, 23245–23263 (1993).
- Powell, M. D. Evaluations of diagnostic marine boundary layer models applied to hurricanes. *Mon. Weath. Rev.* **108**, 757–766 (1980).
- Powell, M. D. & Black, P. G. The relationship of hurricane reconnaissance flight-level wind measurements to winds measured by NOAA's oceanic platforms. *J. Wind Eng. Ind. Aerodyn.* **36**, 381–392 (1990).
- Vickery, P. J., Skerlj, P. F., Steckley, A. C. & Twisdale, L. A. Hurricane wind field model for use in hurricane simulations. *J. Struct. Eng. ASCE* **126**, 1203–1222 (2000).
- Jelesnianski, C. P., Chen, J. & Shaffer, W. A. SLOSH: Sea, Lake, and Overland Surges from Hurricanes. *NOAA Tech. Rep. NWS* 48 (1992).
- Blain, C. A., Westerink, J. J. & Luettich, R. A. Jr The influence of domain size on the response characteristics of a hurricane storm surge model. *J. Geophys. Res.* **99**, 18467–18479 (1994).
- Tolman, H. L. *et al.* Development and implementation of wind generated ocean surface wave models at NCEP. *Weath. Forecast.* **17**, 311–333 (2002).
- Paulson, C. A. The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteorol.* **9**, 857–861 (1970).
- Charnock, H. Wind stress on a water surface. *Q. J. R. Met. Soc.* **81**, 639–640 (1955).
- Garratt, J. R. Review of drag coefficients over oceans and continents. *Mon. Weath. Rev.* **104**, 418–442 (1977).
- Donelan, M. A., Dobson, F. W., Smith, S. D. & Anderson, R. J. On the dependence of sea surface roughness on wave development. *J. Phys. Oceanogr.* **23**, 2143–2149 (1993).
- Smith, S. D. Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature. *J. Geophys. Res.* **93**, 15467–15472 (1988).
- Large, W. G. & Pond, S. Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.* **11**, 324–336 (1981).
- Taylor, P. K. & Yelland, M. J. The dependence of sea surface roughness on the height and steepness of the waves. *J. Phys. Oceanogr.* **31**, 572–590 (2001).
- Vickery, P. J. & Skerlj, P. F. Elimination of exposure D along the hurricane coastline in ASCE 7. *J. Struct. Eng.* **126**, 545–549 (2000).
- Palmen, E. & Riehl, H. Budget of angular momentum and energy in tropical cyclones. *J. Meteorol.* **14**, 150–159 (1957).
- Miller, B. I. A study of the filling of Hurricane Donna (1960) over land. *Mon. Weath. Rev.* **92**, 389–406 (1964).
- Hawkins, H. F. & Rubsam, D. T. Hurricane Hilda Part II: Structure and budgets of the hurricane on 1 Oct. 1964. *Mon. Weath. Rev.* **96**, 617–636 (1968).
- Donelan, M. A. & Hui, W. H. *Surface Waves and Fluxes* (eds Geernaert, G. L. & Plant, W. J.) (Kluwer, Dordrecht, 1990).
- Geernaert, G. L., Katsaros, K. B. & Richter, K. Variation of the drag coefficient and its dependence on sea state. *J. Geophys. Res.* **91**, 7667–7679 (1986).
- Smith, S. D. *et al.* Sea surface wind stress and drag coefficients: the HEXOS results. *Bound. Layer Meteorol.* **60**, 109–142 (1992).
- Ancitl, F. & Donelan, M. A. Air–water momentum flux observations over shoaling waves. *J. Phys. Oceanogr.* **26**, 1344–1353 (1996).
- Krugermeier, L., Gruenewald, M. & Duncel, M. The influence of sea waves on the wind profile. *Bound. Layer Meteorol.* **14**, 403–414 (1978).
- Jansen, P. A. E. M. Wave induced stress and the drag of air flow over sea waves. *J. Phys. Oceanogr.* **19**, 745–754 (1989).
- Large, W. G., Morzel, J. & Crawford, G. B. Accounting for surface wave distortion of the marine wind profile in low-level ocean storms wind measurements. *J. Phys. Oceanogr.* **11**, 2959–2971 (1995).
- Wright, C. W. *et al.* Hurricane directional wave spectrum spatial variation in the open ocean. *J. Phys. Oceanogr.* **31**, 2472–2488 (2001).
- Tannenhill, I. R. *Hurricanes* (Princeton Univ. Press, Princeton, 1944).
- Hock, T. R. & Franklin, J. L. The NCAR GPS dropwindsonde. *Bull. Am. Meteorol. Soc.* **80**, 407–420 (1999).
- Houston, S. H. *et al.* in *Preprints 24th Conf. Hurricanes Tropical Meteorol.* (Fort Lauderdale, Florida, May 29–June 2) 339 (American Meteorological Society, Boston, 2000).
- Willoughby, H. E. Gradient balance in tropical cyclones. *J. Atmos. Sci.* **47**, 265–274 (1990).
- Shaw, N. The birth and death of cyclones. *Geophys. Mem.* **2**, 213–227 (1922).
- Haurwitz, B. The height of tropical cyclones and of the 'eye' of the storm. *Mon. Weath. Rev.* **63**, 45–49 (1935).
- Jorgensen, D. P. Mesoscale and convective scale characteristics of mature hurricanes. II: Inner core structure of Hurricane Allen (1980). *J. Atmos. Sci.* **41**, 1287–1311 (1984).
- Krayer, W. R. & Marshall, R. D. Gust factors applied to hurricane winds. *Bull. Am. Soc.* **73**, 613–617 (1992).
- Vickery, P. J. & Skerlj, P. F. Hurricane gust factors revisited. *J. Struct. Eng.* (submitted).
- Bradbury, W. M. S., Deaves, D. M., Hunt, J. C. R., Kershaw, R. & Nakamura, K. The importance of convective gusts. *Meteorol. Appl.* **1**, 365–378 (1994).
- Powell, M. D., Dodge, P. & Black, M. L. The landfall of Hurricane Hugo in the Carolinas. *Weath. Forecast.* **6**, 379–399 (1991).
- Powell, M. D., Reinhold, T. A. & Marshall, R. D. in *Proc. 10th Int. Conf. Wind Eng. (Copenhagen, 21–24 June)* (eds Larsen, A., Larose, G. L. & Livesey, F. M.) 307–314 (Balkema, Rotterdam, 1999).
- Amorochio, J. & DeVries, J. A new evaluation of the wind stress coefficient over water surfaces. *J. Geophys. Res.* **85**, 433–442 (1980).
- Alamaro, M., Emanuel, K., Cotton, J., McGillis, W. & Edson, J. in *Preprints 25th Conf. Hurricanes Tropical Meteorol.* (San Diego, California, April 29–May 3) 667 (American Meteorological Society, Boston, 2002).
- Newell, A. & Zakharov, V. E. Rough sea foam. *Phys. Rev. Lett.* **69**, 1149–1151 (1992).
- Lundquist, J. in *Abstracts Limnology and Oceanography: Navigating into the Next Century (Santa Fe, New Mexico, February 1–5)* Abstr. S554WE1538S (American Society of Limnology and Oceanography, Waco, Texas, 1999).
- Kepert, J. D., Fairall, C. W. & Bao, J. W. in *Air–Sea Fluxes: Momentum, Heat, and Mass Exchange* (ed. Geernaert, G. L.) 363–409 (Kluwer, Dordrecht, 1998).
- Donnelly, W. J. *et al.* Revised ocean backscatter models at C and Ku band under high-wind conditions. *J. Geophys. Res.* **104**, 11485–11497 (1999).
- Kaplan, J. & Frank, W. M. The large scale inflow-layer structure of Hurricane Frederic (1979). *Mon. Weath. Rev.* **121**, 3–20 (1993).
- Shay, L. K. Upper ocean response to tropical cyclones. *RSMAS Tech. Note 99-003* (Univ. Miami, Rosenstiel School of Marine and Atmospheric Science, Florida, 1999).
- Katsaros, K. B., Vachon, P. W., Liu, W. T., Black, P. G. *J. Oceanography* **58**, 137–151 (2002).
- Wroe, D. R. & Barnes, G. M. Inflow layer energetics of Hurricane Bonnie (1998) near landfall. *Mon. Weath. Rev.* **131** (in the press, 2003).

Acknowledgements This paper is dedicated to the memory of R. Marshall. The assistance of F. Marks, the scientific and support staff of the NOAA Hurricane Research Division in Miami, and the NOAA Aircraft Operations Center in Tampa is appreciated. W. McGillis provided insight on the possible effect of sea foam on momentum transfer in tropical cyclones. Comments on an earlier version of the manuscript by E. Uhlhorn, F. Marks and R. Rogers are appreciated.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P. (e-mail: Mark.Powell@noaa.gov).

Early optical emission from the γ -ray burst of 4 October 2002

D. W. Fox*, S. Yost†, S. R. Kulkarni*, K. Torii‡, T. Kato§, H. Yamaoka||, M. Sako¶, F. A. Harrison†, R. Sari¶, P. A. Price#, E. Berger*, A. M. Soderberg*, S. G. Djorgovski*, A. J. Barth*, S. H. Pravdo☆, D. A. Frail**, A. Gal-Yam††, Y. Lipkin‡‡, T. Mauch‡‡, C. Harrison# & H. Battersby§§

* Caltech Optical Observatories 105-24, † Space Radiation Laboratory 220-47, ¶ Theoretical Astrophysics 130-33, California Institute of Technology (Caltech), Pasadena, California 91125, USA
‡ Cosmic Radiation Laboratory, RIKEN, Wako, Saitama 351-0198, Japan
§ Department of Astronomy, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
|| Department of Physics, Kyushu University, 4-2-1 Ropponmatsu Chuo-ku, Fukuoka 810-8560, Japan
RSAA, Mt Stromlo Observatory, via Cotter Rd, Weston, Australian Capital Territory, 2611, Australia
☆ Jet Propulsion Laboratory, MS 306-438, 4800 Oak Grove Drive, Pasadena, California 91109, USA
** National Radio Astronomy Observatory, PO Box O, Socorro, New Mexico 87801, USA
†† School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel
‡‡ School of Physics, University of Sydney, New South Wales 2006, Australia
§§ Astrophysics Group, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Observations of the long-lived emission—or ‘afterglow’—of long-duration γ -ray bursts place them at cosmological distances, but the origin of these energetic explosions remains a mystery. Observations of optical emission contemporaneous with the burst of γ -rays should provide insight into the details of the explosion, as well as into the structure of the surrounding environment. One bright optical flash was detected during a burst¹, but other efforts^{2,3} have produced negative results. Here we report the discovery of the optical counterpart of GRB021004 only 193 seconds after the event. The initial decline is unexpectedly slow and requires varying energy content in the γ -ray burst blastwave over the course of the first hour. Further analysis of the X-ray and optical afterglow suggests additional energy variations over the first few days.

The HETE-II spacecraft γ -ray burst (GRB) mission triggered on 4 October 2002 at 12:06:13.6 UT (ref. 4), and within minutes we began observations of the 30-arcmin radius error circle with the Palomar 48-inch Oschin Telescope (P48) and the Near Earth Asteroid Tracking (NEAT) camera. When we compared our first image, obtained 537 s after the burst, with subsequent images at 721 s and 1,028 s post-burst and with Palomar Digital Sky Survey (DPOSS) images, a new fading source was revealed about 100 times brighter than the faintest sources in DPOSS; see Fig. 1. We promptly disseminated⁵ its position and magnitude via the GRB Coordinates Network (GCN; see <http://gcn.gsfc.nasa.gov>), proposing it as the probable GRB counterpart.

Separately, the Automated Response Telescopes at The Institute for Physical and Chemical Research (RIKEN) located at Wako, Japan, responded to the trigger in 193 s. Co-added images reveal a source at the position of the P48 counterpart (Fig. 1). Observations over the next 2 h from a 30-cm telescope⁶ located on the Kyoto University campus, Japan, and a 1.01-m telescope⁷ at Bisei Observatory, Japan, found that the source faded. As dawn set in at Palomar, we continued imaging observations at the Siding Spring Observatory (SSO), Australia, and the Wise Observatory, Israel (Table 1).

At the same time we observed the transient with the Double-Beam Spectrograph on the SSO 2.3-m telescope. A 1,200-s spectrum covering 5,450–9,000 Å revealed two distinct absorption-line systems,

at $z = 1.38$ and 1.60, leading us to conclude that the GRB redshift $z \geq 1.60$ (ref. 8). Higher-quality Keck spectroscopy revealed additional features; a 4,050-Å emission line with a P Cygni profile and a blueward continuum decrement, identified⁹ as Lyman α emission from a $z = 2.32$ galaxy and intergalactic (Lyman break) absorption of the afterglow at the same redshift. For this reason, the redshift of GRB021004 is taken to be 2.32.

The brilliance and early detection of the afterglow motivated observations by a large number of ground-based telescopes (by our count, 40 optical/infrared telescopes and six radio telescopes) around the world, resulting in high-quality, densely sampled light curves in band passes from the blue into the near-infrared and radio¹⁰. In Fig. 2 we display the R-band light curve.

We successfully sought a Director's Discretionary Time observation with the Chandra X-ray Observatory. Beginning at 20.5 h we observed the fading source for a total of 88.1 ks with the Advanced CCD Imaging Spectrometer X-ray detector at the focal plane and the High Energy Transmission Grating Spectrometer behind the mirror assembly. The mean observed flux in the 2–10 keV band was $4.3 \times 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$. Detailed examination of the time-integrated spectrum revealed no narrow emission features, with 90%-confidence upper limits on the flux of such features being $\sim 3 \times 10^{-6} \text{ photons cm}^{-2} \text{ s}^{-1}$ for $\lambda \leq 6 \text{ Å}$, and $\sim 1 \times 10^{-6} \text{ photons cm}^{-2} \text{ s}^{-1}$ for $\lambda \geq 6 \text{ Å}$. The spectrum is well fitted by a power law with photon index $\beta_X = 2.1 \pm 0.1$ for an assumed Galactic hydrogen-equivalent column density, $N_H = 1.1 \times 10^{20} \text{ cm}^{-2}$. The 90%-confidence upper limit on the column density at $z = 2.32$ is $1.1 \times 10^{21} \text{ cm}^{-2}$. We broke the data into four equally spaced time bins and found no evidence for changes in N_H or β_X over the course of the observation.

Despite intensive searches^{2,3}, GRB021004 is only the second GRB from which prompt optical emission has been detected. The first was GRB990123 (ref. 1), which is a typical long-duration event ($T_{\text{GRB}} \approx 100 \text{ s}$) with a typical redshift, $z = 1.60$ (ref. 11), but had a

Table 1 Photometry of the optical afterglow of GRB021004 on 4 October 2002

Telescope	Observer	(Mean) UT	mag	Δmag	Band
RIKEN	Auto	12:11:40	15.45	0.17	NF
P48	D.W.F.	12:15:41	15.543	0.044	NF
P48	D.W.F.	12:18:15	15.594	0.068	NF
P48	D.W.F.	12:23:22	15.886	0.059	NF
RIKEN	Auto	12:47:05	16.39	0.18	NF
SSO40	F.A.H.	16:00:19	17.978	0.054	B
SSO40	F.A.H.	16:12:16	16.540	0.019	I
Wise	Y.L.	17:58:35	17.599	0.156	R
Wise	Y.L.	18:07:39	17.146	0.141	I
Wise	Y.L.	18:26:16	16.397	0.061	I
Wise	Y.L.	18:34:15	17.720	0.111	R
Wise	Y.L.	18:43:34	18.138	0.140	V
Wise	Y.L.	18:55:00	19.036	0.263	B
Wise	Y.L.	19:06:10	18.729	0.185	B
Wise	Y.L.	19:16:15	17.945	0.125	R
Wise	Y.L.	19:23:14	18.051	0.131	R
Wise	Y.L.	19:30:12	17.966	0.118	R
Wise	Y.L.	19:37:31	18.132	0.130	R
Wise	Y.L.	19:45:39	17.892	0.110	R
Wise	Y.L.	19:53:37	18.068	0.113	R

The epoch of the GRB trigger is 12:06:14 UT, 4 October 2002 (ref. 4). The P48/NEAT and RIKEN data were obtained without filter (‘NF’). The integration times for the first set of images are 180 s for the 12:11:40 UT RIKEN point, 60 s for the P48 points, and 2,340 s for the 12:47:05 UT RIKEN point (coverage from 12:14:02 to 13:19:37 UT). P48 images were dark-subtracted using a dark exposure taken at the start of the night, and flattened using an object-masked median-filtered flat field prepared from our database of prior NEAT observations. The P48 data have been converted to R-band assuming that the afterglow has the same colour as the median colour of stars in the field²¹, after iterative rejection of outliers in the zero-point fit. Comparison of these results with results obtained from a subset of 16 calibrated stars with blue colours indicates negligible systematic effects from this procedure. The RIKEN photometry was done with respect to a single comparison star GSC1183.403 located at $\alpha = 00:26:57.6$ and $\delta = 18:59:50.8$ (J2000) with approximate magnitude $R = 11.34$ mag, and checked against three $R \leq 14$ calibrated stars from ref. 21. The 12:11:40 UT RIKEN point is derived from a co-added image of six images, and the 12:47:05 UT RIKEN point is derived from a co-added image of 78 images that is nearly identical to the image shown in Fig. 1. SSO and Wise observations were made through standard astronomical filters, reduced by standard procedures, with zero points determined using secondary calibrators in the field²¹.

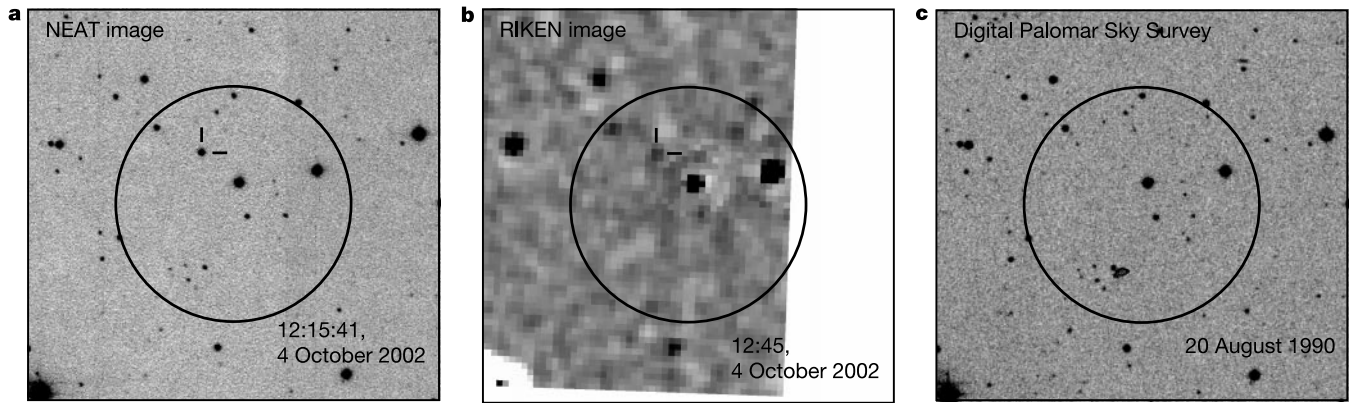


Figure 1 Optical images of GRB021004. These were obtained with the NEAT camera on the P48 telescope (a), and the RIKEN 20-cm Automated Response Telescope (b); the R-band (f -emulsion) DPOSS image is shown for comparison (c). The afterglow is marked by two dashes. The position of the afterglow with respect to the Guide Star Catalog II is $\alpha = 00:26:54.69$ and $\delta = +18:55:41.3$ (J2000) with estimated uncertainty of 0.2 arcsec in each axis. The initial HETE-II localization (radius, 30 arcmin), although large, was encompassed by the large field of view of the NEAT camera, $1.1^\circ \times 1.1^\circ$. The best HETE-II localization at the time of our discovery announcement is indicated by the circle, of radius 2 arcmin. The RIKEN system consists of three equatorial mounts set on top of RIKEN's main building in the city of Wako, Saitama Prefecture, Japan. These mounts respond automatically to email alerts from the GCN. Currently, a 20-cm $f/4.0$ newtonian telescope (T1) and a 25-cm $f/6.8$ reflector (T2) are

mounted on two of the mounts and are equipped with Apogee AP7p 512²-pixel charge-coupled device (CCD) and Apogee AP6E 1024²-pixel CCD, respectively. The integration times for the CCDs are 30 s and 60 s respectively. The fields of view are about 50 arcmin and the observations are obtained without any filter. On 4 October 2002 in response to the HETE-II email alert, T1 and T2 slewed to the GRB position. The first image was obtained starting at 193 s and 209 s for T1 and T2 respectively. The GRB position was close to the edge of the chip on T1. The data from T2 were worse than T1 owing to a focusing problem. The observations were affected by passing clouds as well as significant light pollution from Tokyo. For display purposes we have stacked all of our early time data, 93 images from T1 (mean epoch, 12:45 ut), after convolving with the point spread function.

large fluence¹², $S_{\gamma}(40\text{--}700\text{ keV}) \approx 3 \times 10^{-4} \text{ erg cm}^{-2}$. GRB021004 is a typical long-duration event but is considerably fainter¹³, with fluence $S_{\gamma}(50\text{--}300\text{ keV}) = 1.8 \times 10^{-6} \text{ erg cm}^{-2}$ and total isotropic-equivalent γ -ray energy $E_{\text{iso},\gamma} \approx 5.3 \times 10^{52} \text{ erg}$.

In the framework of the popular 'fireball' model, the early optical emission (by which we mean emission during the burst and on timescales of T_{GRB}) arises as a reverse shock ploughs back through the relativistic ejecta (for example see ref. 14). The same ejecta expand into the circumburst medium, the shock of which produces the long-lived afterglow emission. Observations occurring at time t probe length scales $r \approx 2\gamma^2 ct$, where γ is the Lorentz factor of the blast wave. Early optical observations therefore provide unique diagnostics of the circumburst medium on much smaller length scales than afforded by afterglow observations. The circumburst medium may be either homogeneous (the interstellar medium or ISM model), or may be enriched by the progenitor (if it is a massive star, strong stellar winds produce an inhomogeneous medium with density at radius r , $\rho(r) \propto r^{-2}$; the 'Wind' assumption).

In both models, one expects the reverse shock emission to peak at T_{GRB} and then decay, $f(t) \propto t^{\alpha_R}$ with $\alpha_R \approx -2$ for the ISM model; an even steeper decay is expected for the wind case¹⁵. The optical emission from GRB990123 (ref. 1) is a textbook example of an impulsive explosion taking place in the ISM.

By contrast, the early decay of optical emission from GRB021004 is unexpectedly gradual: on timescales of 200–1,800 s (Fig. 2) we obtain $\alpha_R = -0.4 \pm 0.1$ for GRB021004. This slow decay is not compatible with the usual assumption of a single ejection of a uniform shell. Thus, contrary to ref. 16, our data do not support the standard models for GRB021004. Substantial energy input is required to power this slow decay. Possible explanations include a 'patchy' ejecta shell¹⁷ (with the energy varying as different regions of the shell come into view) or energy injection that continues well beyond T_{GRB} .

GRB021004 was peculiar in another respect, namely, the afterglow emission exhibited strong temporal variation compared to the

canonical power-law decay. Such variations could be produced by circumburst density variations^{17–19} or by energy variations (as discussed above). In principle, the two alternatives can be distinguished¹⁷ by exploiting the fact that the flux $f(\nu > \nu_c)$ at frequencies

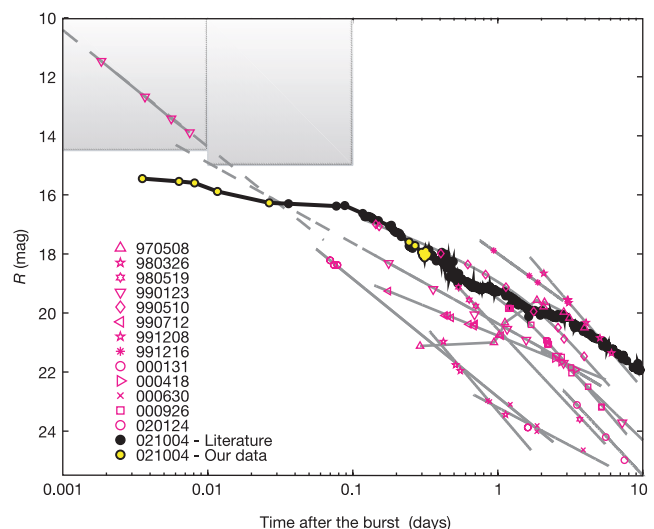


Figure 2 Optical light curve of the afterglow of GRB021004 (refs 6, 22–24) alongside other well-studied afterglows taken from the literature. The R-band data reported here are shown in yellow. The solid lines are the best-fit power-law models, $f(t) \propto t^{\alpha}$. In the case of GRB021004 the solid line simply connects the observations since there is no simple power-law that describes the evolution of the light curve (see also GRB970508). Two distinct power-law fits indicate separately the prompt and afterglow emission of GRB990123. In all cases we have corrected the observed magnitudes for Galactic extinction, but have not applied any correction for extinction within the GRB host galaxy. The shaded box represents the typical limits of detection of experiments such as LOTIS² and ROTSE³.

above the cooling frequency (electrons radiating photons above ν_c lose their energy rapidly) is insensitive to variations in the ambient density (n) whereas below the cooling frequency $f(\nu < \nu_c) \propto n^{1/2}$.

Modelling of the X-ray spectrum and optical data suggest that ν_c is below the X-ray band. Over the period covered by the X-ray observations, the R-band optical flux and the X-ray flux each show significant deviations from a power-law decline (Fig. 3). Subtracting a fitted power-law decay from both data sets, we find that the Spearman's rank correlation coefficient for the X-ray flux deviations (measured at hourly intervals) against the R-band flux deviations (as linearly interpolated to the hourly X-ray epochs) is $r = 0.37$. This indicates correlation at 93% confidence, implying that the varying energy scenario is the more likely explanation for the variations seen during this time frame.

A 93% confidence level is, however, far from conclusive, and the

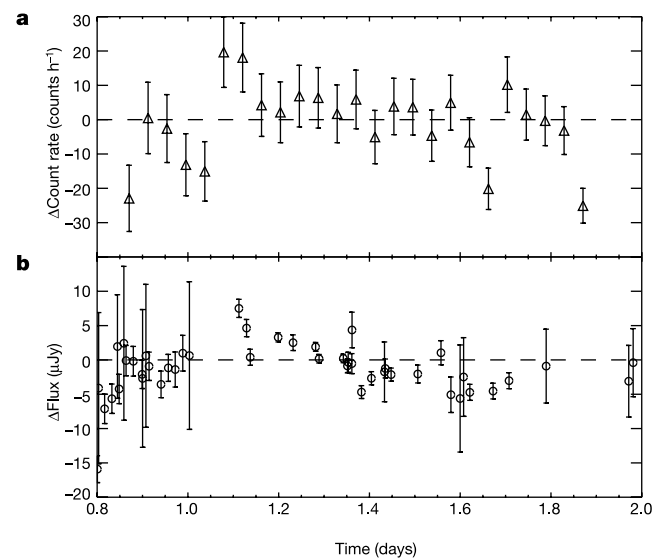


Figure 3 Direct comparison of the X-ray and optical light curves of the afterglow of GRB021004. **a**, The Chandra X-ray observations began at 08:57 UT on 5 October 2002 and ended at 09:26 UT on 6 October 2002. For the Chandra data, we followed standard data screening and analysis procedure (CIAO version 2.2). The dispersed first-order spectra from the Medium and High Energy Gratings were extracted by first applying a 2-arcsec spatial filter along the cross-dispersion direction, and then filtered in CCD pulse-height versus dispersion coordinate space. The spectral response including the effect on low energy response due to quantum efficiency degradation was accounted for by the procedure described at http://asc.harvard.edu/cal/Acis/Cal_prods/qeDeg/. The average count rate for the observation was 0.018 counts s⁻¹ (dispersed spectrum) and 0.014 counts s⁻¹ (zero-order). We produced an exposure-corrected light curve of the X-ray afterglow by binning all events in the 0.5–10 keV band in 3,600 s intervals, correcting for detector live time (which varies by <3% over the observation). The X-ray count rate (in counts h⁻¹) is modelled as $\ln(c_X) = (4.439 \pm 0.037) - (1.195 \pm 0.112) \times \ln(t_d)$ and the fit has $\chi^2 = 50.3$ for 23 degrees of freedom (DOF). **b**, The optical (R-band) data are drawn from Table 1 and the literature (see text) and have not been corrected for Galactic interstellar extinction. To emphasize possible substructure in the light curves we plot the residuals after fitting the data to a decaying power-law model. The observed R-band flux (in μ Jy) is modelled as $\ln(f_R) = (4.223 \pm 0.003) - (1.018 \pm 0.011) \times \ln(t_d)$ and the fit has $\chi^2 = 877$ for 72 DOF. Here, t_d is the time since the burst in days and the above fits are restricted to the time interval $t_d = 0.8$ to 2.0. The large χ^2 values and patterns in the residuals clearly indicate that neither the X-ray nor the optical afterglows exhibit a simple power-law decay over this interval. Making a linear interpolation of the R-band residuals to the evenly-spaced X-ray epochs, we find that Spearman's rank correlation test indicates correlation of the two data sets with 93% confidence, suggesting that energy rather than density variations are responsible for the observed fluctuations.

observed level of correlation may be accidental. Moreover, the rationale discussed above to distinguish density variations from energy injection is strictly valid only in the asymptotic limit. Density fluctuations may in fact cause flux variations above ν_c ; these are expected to be mild in amplitude (we estimate $\lesssim 30\%$, based on ref. 20), but so are the variations we observe. In the future, the X-ray light curves from early times provided by the *Swift* satellite will allow us to distinguish the two cases.

Note added in proof: Since the submission of this Letter, optical emission on even shorter timescales^{25–27} has been observed from the afterglow of GRB021211²⁸. The early emission for this GRB decays rapidly, almost identically to that for GRB990123¹. □

Received 12 December 2002; accepted 19 February 2003; doi:10.1038/nature01504.

1. Akerlof, C. *et al.* Observation of contemporaneous optical radiation from a gamma-ray burst. *Nature* **398**, 400–402 (1999).
2. Williams, G. G. *et al.* LOTIS search for early-time optical afterglows: GRB 971227. *Astrophys. J.* **519**, L25–L29 (1999).
3. Kehoe, R. *et al.* A search for early optical emission from short- and long-duration gamma-ray bursts. *Astrophys. J.* **554**, L159–L162 (2001).
4. Shirasaki, Y. *et al.* GRB021004 (= H2380): a long GRB localized by HETE in near-real time. *GRB Circ. New.* **1565**, 1 (2002).
5. Fox, D. W. GRB021004: optical afterglow. *GRB Circ. New.* **1564**, 1 (2002).
6. Uemura, M., Ishioka, R., Kato, T. & Yamaoka, H. GRB 021004: Kyoto observation. *GRB Circ. New.* **1566**, 1 (2002).
7. Matsumoto, K., Kawabata, T., Ayani, K., Urata, Y. & Yamaoka, H. GRB021004: optical observations at Bisei. *GRB Circ. New.* **1567**, 1 (2002).
8. Fox, D. W. *et al.* GRB021004: absorption redshift. *GRB Circ. New.* **1569**, 1 (2002).
9. Chornock, R. & Filippenko, A. V. GRB 021004 redshift. *GRB Circ. New.* **1605**, 1 (2002).
10. Berger, E., Kulkarni, S. R. & Frail, D. A. GRB 021004: Unusual radio-to-submm spectrum. *GRB Circ. New.* **1612**, 1 (2002).
11. Kulkarni, S. R. *et al.* The afterglow, redshift and extreme energetics of the gamma-ray burst of 23 January 1999. *Nature* **398**, 389–394 (1999).
12. Briggs, M. S. *et al.* Observations of GRB 990123 by the Compton Gamma Ray Observatory. *Astrophys. J.* **524**, 82–91 (1999).
13. Lamb, D. *et al.* HETE fluences for GRB021004. *GRB Circ. New.* **1600**, 1 (2002).
14. Sari, R. & Piran, T. Predictions for the very early afterglow and the optical flash. *Astrophys. J.* **520**, 641–649 (1999).
15. Chevalier, R. A. & Li, Z. Wind interaction models for gamma-ray burst afterglows: The case for two types of progenitors. *Astrophys. J.* **536**, 195–212 (2000).
16. Kobayashi, S. & Zhang, B. GRB021004: Reverse shock emission. *Astrophys. J.* **582**, L75–L78 (2003).
17. Nakar, E., Piran, T. & Granot, J. Variability in GRB Afterglows and GRB 021004. *New Astron.* (submitted); preprint at arXiv.org/astro-ph/0210631 (2002).
18. Lazzati, D., Rossi, E., Covino, S., Ghisellini, G. & Malesani, D. The afterglow of GRB 021004: surfing on density waves. *Astron. Astrophys.* **396**, L5–L9 (2002).
19. Heyl, J. S. & Perna, R. Broadband modeling of GRB 021004. *Astrophys. J. Lett.* (in the press); preprint at arXiv.org/astro-ph/0211256 (2002).
20. Granot, J. & Sari, R. The shape of spectral breaks in gamma-ray burst afterglows. *Astrophys. J.* **568**, 820–829 (2002).
21. Henden, A. GRB021004, BVRC field photometry. *GRB Circ. New.* **1583**, 1 (2002).
22. Holland, S. T. *et al.* Optical photometry of GRB 021004: The first month. *Astrophys. J.* (in the press); preprint at arXiv.org/astro-ph/0211094 (2002).
23. Pandey, S. B. *et al.* Optical observations of the bright long duration peculiar GRB 021004 afterglow. *Bull. Astron. Soc. India* (submitted); preprint at arXiv.org/astro-ph/0211108 (2002).
24. Bersier, D. *et al.* The unusual optical afterglow of the gamma-ray burst GRB 021004: Color changes and short-time-scale variability. *Astrophys. J.* **584**, L43–L46 (2003).
25. Park, H. S., Williams, G. & Barthelmy, S. GRB021211 (HETE2493), OT at 143 sec. *GRB Circ. New.* **1736**, 1 (2002).
26. Wozniak, P. *et al.* GRB021211: Measurement of early time afterglow. *GRB Circ. New.* **1757**, 1 (2002).
27. Li, W., Filippenko, A. V., Chornock, R. & Jha, S. The early light curve of the optical afterglow of GRB021211. Preprint at arXiv.org/astro-ph/0302136 (2003).
28. Fox, D. W. *et al.* Discovery of early optical emission from GRB021211. Preprint at arXiv.org/astro-ph/0301377 (2003).

Acknowledgements We thank D. L. Kaplan and C. Trujillo for assistance with supporting observations at Palomar and R. D. Blandford for discussions. The results presented in this paper are based on data obtained using the NASA-JPL CCD camera at the Oschin Telescope, Palomar Observatory, California Institute of Technology (Caltech). The research was performed in part by the Jet Propulsion Laboratory, Caltech, under contract with the National Aeronautics and Space Administration (NASA). γ -ray burst research at Caltech is supported in part by funds from NSF and NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W.F. (e-mail: derekfox@astro.caltech.edu).

Femtosecond X-ray measurement of coherent lattice vibrations near the Lindemann stability limit

Klaus Sokolowski-Tinten*, Christian Blome*, Juris Blums*, Andrea Cavalleri†, Clemens Dietrich*, Alexander Tarasevitch*, Ingo Uschmann‡, Eckhard Förster‡, Martin Kammler§, Michael Horn-von-Hoegen* & Dietrich von der Linde*

* Institut für Laser- und Plasmaphysik, Universität Essen, 45117 Essen, Germany

† Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

‡ Forschungsgruppe Röntgenoptik, Institut für Optik und Quantenelektronik, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany

§ Institut für Halbleitertechnologie, Universität Hannover, 30167 Hannover, Germany

The study of phase-transition dynamics in solids beyond a time-averaged kinetic description requires direct measurement of the changes in the atomic configuration along the physical pathways leading to the new phase. The timescale of interest is in the range 10^{-14} to 10^{-12} s. Until recently, only optical techniques were capable of providing adequate time resolution¹, albeit with indirect sensitivity to structural arrangement. Ultrafast laser-induced changes of long-range order have recently been directly established for some materials using time-resolved X-ray diffraction^{2–8}. However, the measurement of the atomic displacements within the unit cell, as well as their relationship with the stability limit of a structural phase^{9–11}, has to date remained obscure. Here we report time-resolved X-ray diffraction measurements of the coherent atomic displacement of the lattice atoms in photoexcited bismuth close to a phase transition. Excitation of large-amplitude coherent optical phonons gives rise to a periodic modulation of the X-ray diffraction efficiency. Stronger excitation corresponding to atomic displacements exceeding 10 per cent of the nearest-neighbour distance—near the Lindemann limit—leads to a subsequent loss of long-range order, which is most probably due to melting of the material.

The most common crystalline form of bismuth is the α -arsenic or

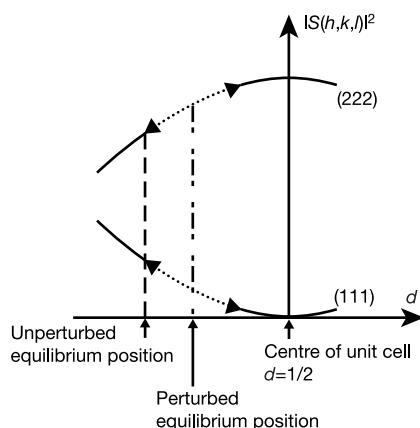


Figure 1 Geometrical structure factor $|S(h, k, l)|^2$ of bismuth as a function of the distance d of the two basis atoms for diffraction from (111) and (222) lattice planes. The dotted lines denote the atomic movement after an impulsive shift of the equilibrium Bi–Bi distance upon transition to an excited electronic state.

A7 structure¹². The lattice can be derived from the face-centred-cubic lattice by a weak rhombohedral distortion of the cubic unit cell. The crystal basis consists of two Bi atoms: atom 1 on a lattice site, and atom 2 at a position slightly displaced from the centre along the body diagonal of the unit cell.

The distortion from the cubic structure to A7 is stabilized by the so-called Peierls–Jones¹³ mechanism, which introduces a small bandgap over extended regions of the Brillouin zone and is responsible for bismuth being a semimetal. However, the equilibrium structure, particularly the equilibrium distance of the two basis atoms, can be easily perturbed by external influences—for example, by changes of pressure and temperature, or by electronic excitation^{14,15}. When the distance is changed by a fast perturbation, the Bi atoms perform a damped oscillation along the body diagonal and finally come to rest at the new equilibrium position. This mode of coherent atomic motion represents a totally symmetric A_{1g} optical phonon mode of the crystal.

Excitation of coherent lattice vibrations by an impulsive perturbation of the atomic equilibrium position (displacive excitation) has previously been studied in a variety of materials^{15–20}, including bismuth. In these experiments femtosecond laser excitation was used, and the atomic motion was detected indirectly by measuring the weak oscillations in the optical reflectivity at the phonon frequency that are produced by the lattice vibrations.

A direct way of observing the evolution of the atomic configuration is to use time-resolved X-ray diffraction. The changes in the diffraction efficiency caused by atomic motion in the unit cell are determined by the squared modulus of the geometrical structure factor $|S(h, k, l)|^2$, where h , k and l are Miller indices. Figure 1 compares the X-ray diffraction from lattice planes of bismuth corresponding to $(hkl) = (111)$ and $(hkl) = (222)$. $|S|^2$ is plotted as a function of the Bi–Bi distance d (in units of the length of the body diagonal). For $d = 1/2$ (atom 2 in the centre), diffraction from (111) planes is forbidden, whereas (222) diffraction has a maximum. The actual equilibrium value of d in the unperturbed crystal is marked by the dashed vertical line ($d = 0.468$).

We now consider a perturbation that sets the atoms in motion towards a new equilibrium position with atom 2 closer to the centre of the unit cell (dash-dotted line in Fig. 1). As this motion proceeds, we expect to see changes in the X-ray diffraction signal in opposite directions: an increase for (222) and a decrease for (111). Following these initial changes, the oscillatory atomic motion should give rise to oscillations both in the (222) and the (111) X-ray diffraction with a frequency corresponding to the A_{1g} optical phonon mode.

In the time-resolved X-ray diffraction experiments reported here,

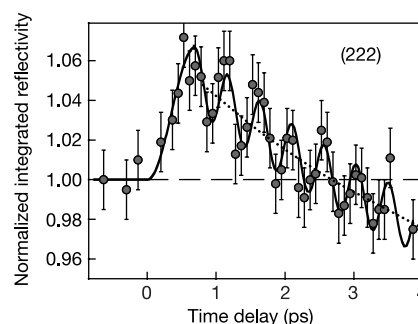


Figure 2 X-ray diffraction efficiency of the (222) reflection as a function of time delay between the optical pump pulse (fluence 6 mJ cm^{-2}) and the X-ray probe pulse. The solid line is fitted to the experimental data. The decrease of the mean value of the X-ray signal (dotted line) is explained by the Debye–Waller effect, and reflects the increasing random component of the atomic motion.

we studied crystalline bismuth films of 50 nm thickness and (111) surface orientation. X-ray pulses at a photon energy of 4.51 keV (wavelength 0.274 nm) were used. The material was optically excited by near-infrared (800 nm) femtosecond laser pulses with an energy fluence of 6 mJ cm^{-2} in order to launch a strong A_{1g} optical phonon mode through displacive excitation. The atomic motion underlying the excited lattice wave was probed by measuring the diffraction of the X-ray probe pulses interacting with the lattice after a controlled time delay with respect to the laser pulses. Measurements were performed at two different Bragg angles, $\theta_{111} = 20^\circ$ and $\theta_{222} = 44^\circ$, in order to compare the X-ray diffraction from (111) and (222) lattice planes.

The result of the (222) experiment is shown in Fig. 2. The measured X-ray diffraction signal (integrated reflectivity) is plotted as a function of the time delay between the laser pulses and the X-ray pulses. As the time is increased, an initial increase in the diffraction signal within a few hundred femtoseconds is observed. The first maximum of the diffraction signal is followed by a sequence of distinct periodic minima and maxima. This feature of the data clearly indicates oscillations in the temporal evolution of the X-ray diffraction efficiency. The solid curve in Fig. 2 is a fit of the experimental data. The oscillatory part of the fit consists of constant cycles of 467 fs, corresponding to a frequency $\nu = 2.12 \text{ THz}$. No effort has been made to take into account damping or changes in the cycle duration.

We performed similar measurements for a crystal orientation in which the diffraction from (111) lattice planes could be recorded. In this case an initial decrease in the diffraction is observed, followed by a series of maxima and minima very similar to what was measured in the (222) diffraction experiment, but in the opposite sense (Fig. 3). The period of these oscillations agrees with the oscillation frequency $\nu = 2.12 \text{ THz}$ observed in the (222) experiment. (The data point at the far right in Fig. 3 represents diffraction measurements taken several seconds after the laser excitation pulse. A full recovery of the crystal diffraction is observed, indicating that no laser damage or other material modifications have occurred.)

From the measured changes in the X-ray diffraction, the actual atomic displacement can be obtained. If it is assumed that the variation of the X-ray signal is dominated by changes in d , the initial increase of 7% in the (222) diffraction corresponds to a Δd of approximately 0.015–0.025 nm. The two principal sources of error in Δd , besides the experimental error bars, are uncertainties concerning the depth profile of the excited Bi layer and the smoothing of the oscillations caused by the finite duration of the X-ray probe pulses. Linear absorption data of Bi correspond to a penetration depth of 14–22 nm (ref. 12), suggesting that only a fraction of the 50-nm Bi film is optically excited and contributes

to changes in the X-ray diffraction signal. From independent experiments we have obtained a value of about 300 fs for the width of the X-ray pulses²¹. Deconvolution calculations using this pulse width imply that the actual oscillation amplitudes are approximately twice as large as the observed oscillations of the X-ray signal.

A comparison of the observed frequency $\nu = 2.12 \text{ THz}$ with the A_{1g} optical phonon frequency of the unperturbed bismuth crystal ($\nu_{1g} = 2.92 \text{ THz}$) indicated that the phonon frequency is significantly downshifted in the highly excited material. Very recently, frequency downshifts to 2.45 THz were observed by Hase *et al.*²⁰ in displacive excitation experiments in which the coherent phonons were detected through changes in the optical reflectivity. The frequency shifts were found to be dependent on the phonon oscillation amplitude, and the effect was attributed to anharmonicity²¹. Here, we observe a much larger downshift. The measured maximum atomic displacements correspond to a significant fraction of the nearest-neighbour distance ($d_{nn} = 0.31 \text{ nm}$) of the Bi atoms in the A7 structure, $\Delta d/d_{nn} \approx 5\text{--}8\%$. Such large atomic displacements are expected to lead to important changes in the crystal potential, and eventually to a destabilization of the crystal structure and structural phase transitions. The Lindemann criterion⁸ states that the solid-to-liquid phase transition occurs when the mean displacement of the lattice atoms reaches 10–20% of the lattice constant.

Indeed, we observed structural disordering of bismuth when the laser excitation was further increased. An example is depicted in Fig. 4, which shows the integrated X-ray reflectivity of the (222) diffraction as a function of pump–probe time delay for an excitation fluence of 20 mJ cm^{-2} . At this excitation level, an initial increase in the diffraction signal as large as 20% within a few hundred femtoseconds was observed. Subject to the uncertainties described above, we conclude that this strong change in the diffraction signal corresponds to an initial coherent atomic displacement of nearly 0.04 nm, or more than 10% of the nearest-neighbour distance.

However, in clear contrast to the observations at lower excitation, no oscillations of the diffraction signal occurred. Instead, it was observed that after reaching the maximum the diffraction signal decreased monotonically, and reached a quasi-stationary level of 40% in approximately 10 ps. We interpret this decrease as an indication of disordering due to melting of the material. This explanation is supported by the observed shift of the Bragg peak towards smaller diffraction angles. The shifts can be attributed to thermal expansion, and indicate that the temperature of the non-disordered material underneath the disordered surface layer is close to the melting point of bismuth. Diffraction data on the (111) reflection are in line with this interpretation.

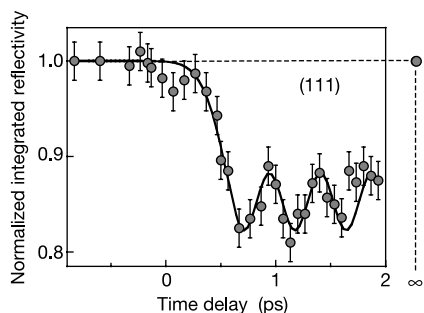


Figure 3 X-ray diffraction efficiency of the (111) reflection as a function of time delay between the optical pump pulse (fluence 6 mJ cm^{-2}) and the X-ray probe pulse. The solid line is fitted to the experimental data. The Debye–Waller effect in (111) diffraction is negligible.

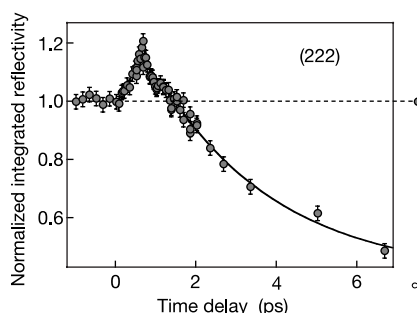


Figure 4 X-ray diffraction efficiency of the (222) reflection as a function of time delay between the optical pump pulse (fluence 20 mJ cm^{-2}) and the X-ray probe pulse.

We believe that the present study represents a significant step beyond the observation of long-range order loss that was made in previous X-ray experiments on ultrafast non-thermal melting: in these previous experiments, it was not possible to obtain the initial displacements. Although the relationship between the large coherent displacements of the lattice atoms observed in the present work and the stability of the parent phase is still elusive, our experiments establish a direction for future studies—the use of X-ray scattering on the appropriate timescale to reveal details of product phase formation. □

Methods

The X-ray pulses were generated by focusing femtosecond laser pulses from a titanium:sapphire laser system (120-fs pulse width, about 100-mJ pulse energy, 10-Hz repetition rate) onto a thin titanium wire to produce a microplasma²¹. The incoherent X-ray emission from the laser-produced plasma (Ti K α line at 0.274 nm) was collected, and focused onto a spot of approximately 80 μ m in the centre of the laser-excited area on the surface of the bismuth crystal. To focus the X-rays, we used a toroidally bent silicon crystal. The 50-nm Bi films were grown heteroepitaxially on a Si(111) wafer. The growth process provided high-quality single crystalline layers of bismuth with a lattice constant different from that of the Si substrate. This allows us to isolate the relatively weak X-ray diffraction of the thin Bi film from the strong background diffraction due to the Si substrate. An X-ray-sensitive electronic camera was used to record the angular distribution of the diffracted X-rays. The X-ray signal integrated over the diffraction angle (integrated reflectivity) typically corresponded to 10 diffracted photons per pulse. Therefore, signal integration over many thousands of pulses was necessary to achieve a measuring accuracy of several per cent.

Received 7 October 2002; accepted 7 February 2003; doi:10.1038/nature01490.

1. Bloembergen, N. From nanosecond to femtosecond science. *Rev. Mod. Phys.* **71**, S283–S287 (1999).
2. Chin, A. H. *et al.* Ultrafast structural dynamics in InSb probed by time-resolved X-ray diffraction. *Phys. Rev. Lett.* **83**, 336–339 (1999).
3. Siders, C. W. *et al.* Detection of non-thermal melting by ultrafast X-ray diffraction. *Science* **286**, 1340–1342 (1999).
4. Lindenberg, A. M. *et al.* Time-resolved X-ray diffraction from coherent phonons during a laser-induced phase transition. *Phys. Rev. Lett.* **84**, 111–114 (2000).
5. Rousse, A. *et al.* Non-thermal melting in semiconductors measured at femtosecond resolution. *Nature* **410**, 65–68 (2001).
6. Sokolowski-Tinten, K. *et al.* Femtosecond X-ray measurement of ultrafast melting and large acoustic transients. *Phys. Rev. Lett.* **87**, 225701 (2001).
7. Cavalleri, A. *et al.* Femtosecond structural dynamics in VO₂ during an ultrafast solid-solid phase transition. *Phys. Rev. Lett.* **87**, 237401 (2001).
8. Feuer, T. *et al.* Femtosecond silicon K α pulses from laser produced plasmas. *Phys. Rev. E* **65**, 016412 (2002).
9. Lindemann, F. A. Über die Berechnung molekularer Eigenfrequenzen. *Phys. Z.* **11**, 609–612 (1910).
10. Born, M. Thermodynamics of crystals and melting. *J. Chem. Phys.* **7**, 591–603 (1939).
11. Tallon, J. L. A hierarchy of catastrophes as a succession of stability limits for the crystalline state. *Nature* **342**, 658–660 (1989).
12. Madelung, O. (ed.) *Semiconductors: Physics of Non-tetrahedrally Bonded Elements and Binary Compounds I* in Landolt-Börnstein, New Series, Group III: *Crystal and Solid State Physics* Vol. 17, *Semiconductors*, Part a (Springer, Berlin, 1983).
13. Peierls, R. *More Surprises in Theoretical Physics* (Princeton Univ. Press, Princeton, 1991).
14. Shick, A. B., Ketterson, J. B., Novikov, D. L. & Freeman, A. J. Electronic structure, phase stability, and semimetal-semiconductor transitions in Bi. *Phys. Rev. B* **60**, 15484–15487 (1999).
15. Zeiger, H. J. *et al.* Theory for dispersive excitation of coherent phonons. *Phys. Rev. B* **45**, 768–778 (1992).
16. Hunsche, S., Wienecke, K., Dekorsy, T. & Kurz, H. Impulsive softening of coherent phonons in tellurium. *Phys. Rev. Lett.* **75**, 1815–1818 (1995).
17. Garrett, G. A., Albrecht, T. F., Whitaker, J. F. & Merlin, R. Coherent THz phonons driven by light pulses and the Sb problem: what is the mechanism? *Phys. Rev. Lett.* **77**, 3661–3664 (1996).
18. Hase, M., Mizoguchi, K., Harima, H. & Nakashima, S. Dynamics of coherent phonons in bismuth generated by ultrashort light pulses. *Phys. Rev. B* **58**, 5448–5452 (1998).
19. DeCamp, M. F., Reis, D. A., Bucksbaum, P. H. & Merlin, R. Dynamics and coherent control of high-amplitude phonons in bismuth. *Phys. Rev. B* **64**, 092301 (2001).
20. Hase, M., Kitajima, M., Nakashima, S. & Mizoguchi, K. Dynamics of coherent anharmonic phonons in bismuth using high density photoexcitation. *Phys. Rev. Lett.* **88**, 067401 (2002).
21. Von der Linde, D. *et al.* 'Ultrafast' extended to X-rays: Femtosecond time-resolved X-ray diffraction. *Z. Phys. Chem.* **215**, 1527–1541 (2001).

Acknowledgements Financial support by the Deutsche Forschungsgemeinschaft, the European Community (Research and Training Network XPOSE) and the German Academic Exchange Service (DAAD) is acknowledged.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S.-T. (e-mail: kst@ilp.physik.uni-essen.de).

Phonon interpretation of the 'boson peak' in supercooled liquids

T. S. Grigera^{*†}, V. Martin-Mayor^{*†}, G. Parisi^{*} & P. Verrocchio^{†‡}

^{*} Dipartimento di Fisica, Sezione INFN, SMC and INFM unità di Roma 1, Università di Roma "La Sapienza", Piazzale Aldo Moro 2, I-00185 Roma, Italy
[†] Dipartimento di Fisica and INFM unità di Trento, Università di Trento, Via Sommarive, 14, 38050 Povo, Trento, Italy

Glasses^{1,2} are amorphous solids, in the sense that they display elastic behaviour. In crystalline solids, elasticity is associated with phonons, which are quantized vibrational excitations. Phonon-like excitations also exist in glasses at very high (terahertz; 10¹² Hz) frequencies; surprisingly, these persist in the supercooled liquids³. A universal feature of such amorphous systems is the boson peak: the vibrational density of states has an excess compared to the Debye squared-frequency law. Here we investigate the origin of this feature by studying the spectra of inherent structures⁴ (local minima of the potential energy) in a realistic glass model. We claim that the peak is the signature of a phase transition in the space of the stationary points of the energy, from a minima-dominated phase (with phonons) at low energy to a saddle-point-dominated phase^{5–7} (without phonons). The boson peak moves to lower frequencies on approaching the phonon–saddle transition, and its height diverges at the critical point. Our numerical results agree with the predictions of euclidean random matrix theory⁸ on the existence of a sharp phase transition⁹ between an amorphous elastic phase and a phonon-free one.

Vibrational excitations of glasses in the terahertz region, and the related vibrational density of states (VDOS), are important in their thermodynamic properties¹⁰. Recent results suggest that the VDOS is determined by the properties of the potential energy landscape⁴ (PEL), a tool useful in the understanding of the slow dynamics of supercooled liquids². Indeed, the VDOS and the dynamic structure factor can be qualitatively reproduced from the (harmonic) vibrational spectrum obtained from the diagonalization of the hessian matrix of the potential energy evaluated at the minima ('inherent structures') of a Lennard–Jones system¹¹. The same numerical procedure has led to quantitative agreement with inelastic X-ray scattering experiments in amorphous silica¹². Here we will show that a boson peak must be present in the VDOS as a consequence of the PEL geometry; we also make quantitative predictions, which can be checked by rapidly quenching supercooled liquids below their glass temperature.

The high-frequency (0.1–10 THz) excitations have been experimentally shown to have linear dispersion relations^{3,13–18} in the mesoscopic momentum region (~ 1 – 10 nm^{–1}). Although clearly not plane waves, they are highly reminiscent of phonons, both because they propagate with the speed of sound, and because the VDOS $g(\omega)$ (ω is the frequency) is Debye-like ($g(\omega) \propto \omega^2$) at low enough frequency. These excitations have in fact been dubbed 'high-frequency sound', and a coherent theoretical picture of their properties has been obtained^{9,19,20} using euclidean random matrix theory⁸ (ERMT).

However, there is more to the low-frequency VDOS of glasses than the Debye law. A characteristic feature is that the VDOS departs from the Debye form, displaying an excess of states, which has been named the boson peak. As observed in many materials^{3,13–18}, the boson peak is in a region of frequencies where

[†] Present addresses: Centro di Studi e Ricerche "Enrico Fermi", via Panisperna, 89/A, 00184 Roma, Italy (T.S.G.); Departamento de Física Teórica I, Universidad Complutense de Madrid, Avenida Complutense, 28040 Madrid, Spain (V.M.-M., P.V.).

the dispersion relation for phonons is still linear (ω is a few terahertz). Several non-equivalent ways of experimentally defining the boson peak have been introduced. One approach defines it as a peak in Raman scattering data, another as a peak in the difference between the VDOS of the glass and the corresponding crystal. Other authors extract $g(\omega)$ from their neutron scattering data and look for a peak in $g(\omega)/\omega^2$. This is the definition that we adopt, because we believe that it reveals universal properties of glasses.

The origin of the boson peak can be understood if we consider the ensemble of generalized inherent structures (GISs): for each equilibrium configuration, the associated GIS is the nearest stationary point of the hamiltonian. If we start from an equilibrium configuration at low temperature, the GIS is a local minimum, and it coincides with the more frequently used inherent structure (that is, the nearest minimum of the hamiltonian). On the contrary, if we start from high temperature, the GISs are saddle points. In the GIS ensemble there is a sharp phase transition separating these two regimes. It takes place in glass-forming liquids^{5–7} at the mode-coupling temperature^{21,22} (T_{MC}), above which liquid diffusion is no

longer ruled by rare ‘activated’ jumps between inherent structures, but by motion along the unstable directions of saddles. Phonons are present in the spectrum of the VDOS in the low-temperature phase (inherent-structure dominated), but are absent in the saddle phase (see ref. 23 for a different approach).

The key point, following both from analytic computations in soluble models²⁴ and simulations^{5–7}, is that the minima obtained starting from configurations below T_{MC} and the saddles obtained starting above T_{MC} join smoothly at T_{MC} . Thus we can study GISs as a single ensemble, parameterized by the initial temperature or the final energy. We expect that the GIS ensemble belongs to a large universality class.

This transition from the phonon phase to the saddle phase is a general phenomenon that is not restricted to glasses. It has been studied in the framework of ERMT⁸, showing⁹ that close to this transition, a boson peak is present in the phonon phase, whose position shifts to lower frequencies on approaching the transition point, while its height grows without bound. The boson peak actually signals a crossover at frequency ω_{BP} between a (phonon dominated) ω^2 scaling of $g(\omega)$ to an ω^γ scaling ($\gamma < 2$) that is present at the phase transition point. More precisely, at frequencies that are small with respect to the Debye frequency, the VDOS should satisfy the scaling law

$$g(\omega, \Delta) = \omega^\gamma h(\omega \Delta^{-\rho}) \quad (1)$$

where Δ is the distance from the critical point and depends on the actual control parameter (pressure, temperature, and so on) and $h(x)$ is such that $h(x) \propto x^{2-\gamma}$ for $x \ll 1$ and $h(x)$ is approximately equal to a constant for $x \gg 1$. This scaling law implies that

$$\omega_{BP} \propto \Delta^\rho, \quad g(\omega_{BP})/\omega_{BP}^2 \propto \Delta^{-\beta} \quad (2)$$

where $\beta = \rho(2 - \gamma)$. Under the resummation of a given class of diagrams, ERMT predicts⁹ that $\rho = 1$, $\gamma = 3/2$ and $\beta = 1/2$. The actual (universal) values of these critical exponents in three dimensions may differ slightly from the values found in this approximation.

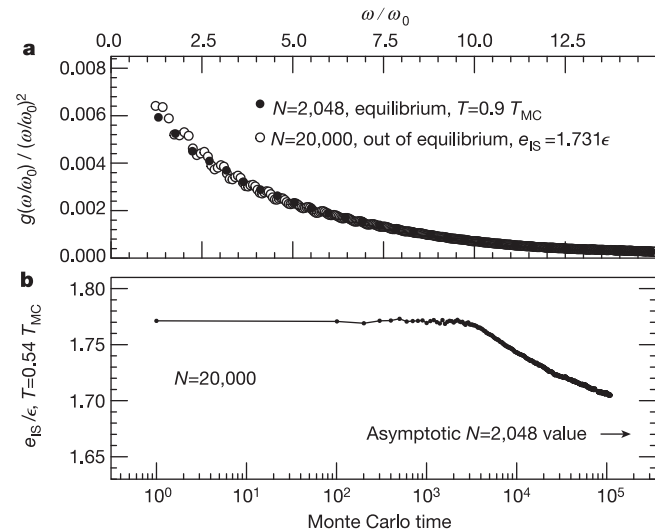


Figure 1 The vibrational density of states, $g(\omega)$, at low frequencies depends only on the energy of the inherent structure, e_{IS} . **a**, $g(\omega/\omega_0)/(\omega/\omega_0)^2$ (where ω_0 is the unit of frequency defined below) for an equilibrated 2,048-particle system at $T = 0.9T_{MC}$ compared to that obtained from ten inherent structures of the same e_{IS} from a 20,000-particle system out of equilibrium. The larger system was quenched from infinite temperature to $T = 0.54T_{MC}$, and the inherent structures obtained by minimizing instantaneous configurations between 1.9×10^4 and 2×10^4 Swap Monte Carlo steps. Similar results are obtained at all the temperatures for which comparable e_{IS} are obtained in the quench. Inherent structures were found by minimizing instantaneous configurations with a conjugate gradient algorithm. The hessian was diagonalized with standard library routines in the case of the smaller system, and with the method of moments for the larger system. Each VDOS was obtained from at least 200 inherent structure as the histogram of the square root of the eigenvalues. **b**, Monte Carlo history of e_{IS} on a logarithmic timescale, for the system of 20,000 particles suddenly quenched from infinite temperature to $T = 0.54T_{MC}$. The soft-sphere binary mixture is made of 50% of particles of type A and 50% of type B, both with the same mass. The interaction potential is $V_{\alpha\beta}(r) = \epsilon[(\sigma_\alpha + \sigma_\beta)/r]^{12} + C_{\alpha\beta}(\alpha, \beta \in \{A, B\})$, with σ values fixed by the conditions $\sigma_B = 1.2\sigma_A$, and $(2\sigma_A)^3 + 2(\sigma_A + \sigma_B)^3 + (2\sigma_B)^3 = 4\sigma_0^3$. A smooth cut-off is imposed at $r_c = \sqrt{3}\sigma_0$: for $r_c \leq r \leq a$, we set $V_{\alpha\beta} = B_{\alpha\beta}(a - r)^3$, and $V_{\alpha\beta} = 0$ for $r > a$. In the results, lengths are given in units of σ_0 , energies in units of ϵ , and frequencies in units of $\omega_0 = (m\sigma_0^2/\epsilon)^{-1/2}$. Using argon parameters ($\sigma_0 = 3.4 \text{ \AA}$, $\epsilon = 120 \text{ K } k_B$ (k_B is Boltzmann's constant), and $m = 39.96 \text{ a.m.u.}$), the frequency unit ω_0 is 0.46 THz and the mode coupling temperature T_{MC} is 26.4 K. All runs are at constant volume, with particle density $\rho = \sigma_0^{-3}$. For temperatures down to $T = 0.9T_{MC}$ thermalization was achieved, as checked by ensuring that the equilibrium fluctuation–dissipation ratio is reached. For $T < 0.9T_{MC}$, the runs were followed until e_{IS} reached stationarity.

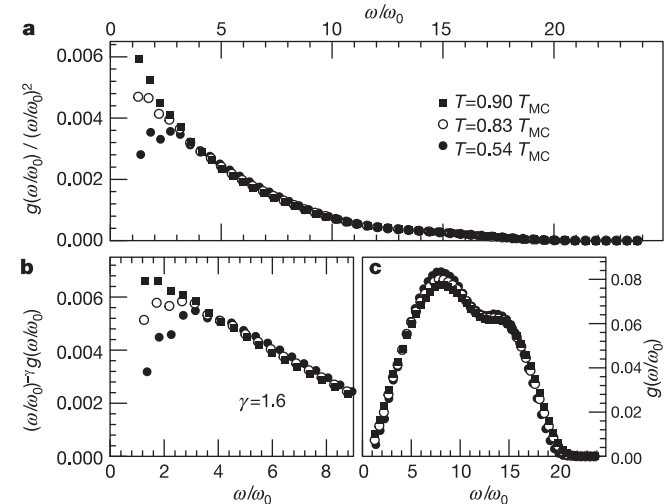


Figure 2 The vibrational density of states $g(\omega)$ of the soft-sphere binary mixture at three representative temperatures. (The full set of temperatures was $T/T_{MC} = 0.9, 0.83, 0.78, 0.69, 0.61, 0.54$ and 0.49 ; frequencies are given in units of ω_0). **a**, The $g(\omega)/\omega^2$ plot shows a boson peak at low temperature (shown for $T = 0.54T_{MC}$). When temperature is increased, the peak shifts to lower frequencies and grows without bound. **b**, The divergence seen in **a** is due to an ω^γ (with $\gamma < 2$) scaling of $g(\omega)$. At temperatures where the boson peak is still seen, a crossover between ω^2 and ω^γ scaling is noted (compare **a** and **b**). Although γ cannot be extracted very accurately from the data, it is compatible with $\gamma = 2 - \beta/\rho \approx 1.6$, with β and ρ taken from the fits of Fig. 3. **c**, The high-frequency features of $g(\omega)$ do not change drastically with temperature.

We now analyse numerically the existence of such a transition by studying the soft-sphere binary mixture, a model of a fragile glass²⁵. We simulate this system with the Swap Monte Carlo algorithm²⁶, and compute the VDOS of the inherent structures obtained starting from equilibrium configurations at temperatures below T_{MC} . Except where stated, we used 2,048 particles, and runs were followed until the energy of the inherent structure, e_{IS} , reached a stationary value. The predictions above agree with the numerical data, taking $\Delta = e_c - e_{IS}$ (where e_c is the critical value of e_{IS}).

Given the numerical results of ref. 27, we expect that $g(\omega)$ will depend only on e_{IS} , and thus that e_{IS} is the relevant control parameter. To check this, we compare the VDOS of two systems at different temperatures (one in equilibrium, the other not) but with the same e_{IS} . The equilibrated system has 2,048 particles, whereas the other has 20,000 and is in the regime where e_{IS} has not yet stabilized after a quench (Fig. 1b). As Fig. 1a shows for a representative case, the VDOS obtained at the same e_{IS} for the two systems coincides, confirming that e_{IS} is indeed the control parameter. We can thus proceed to investigate the scaling laws, shown in equation (2).

The VDOS at three representative temperatures is shown in Fig. 2. For temperatures up to $T = 0.69T_{MC}$, a Debye behaviour ($g(\omega) \propto \omega^2$) is found for the lowest frequencies that can be resolved in a system of this size. In the $g(\omega)/\omega^2$ plot (Fig. 2a), a peak is clearly identified, which is seen to grow in height and shift to lower frequency on raising the temperature. The peak shifts to frequencies below our resolution for $T > 0.69T_{MC}$, but we can identify a Debye region up to $T = 0.83T_{MC}$. Above this temperature the Debye region is not well defined, and it disappears completely for $T \geq 0.9T_{MC}$. At these temperatures, $g(\omega)/\omega^2$ seems to diverge (at least within the frequency range that we can resolve). Experimentally, this behaviour may be difficult to observe owing to the presence of the elastic peak, but B_2O_3 data²⁸ seem to support this prediction. For frequencies immediately above ω_{BP} , the VDOS scales as ω^γ , with $\gamma \approx 1.5$ (Fig. 2b). Note that these changes of the VDOS with temperature go practically unnoticed if one plots $g(\omega)$ directly (Fig. 2c). In particular, the boson peak is completely

unrelated to the first maximum of $g(\omega)$, whose position is insensitive to temperature changes.

Because we know that e_{IS} decreases slowly with time after a quench²⁷ (see also Fig. 1b), the above results enable us to make predictions about the ageing of the VDOS, and in particular of the boson peak. With increasing time the system moves farther from the critical point, and thus the boson peak should decrease in height and shift to higher frequencies. Moreover, at a given frequency (below the asymptotic ω_{BP}) $g(\omega)$ should decrease, as it will be of the order of ω^γ at short times and of the order of ω^2 at very long times. Similarly, we expect a cooling-rate dependence of the shape of the boson peak: the slower the cooling, the lower the asymptotic e_{IS} , and thus the larger the ω_{BP} and the less pronounced the boson peak. An effect of this kind has been observed in As_3S_3 (ref. 29), although in that work it is difficult to disentangle the physical effect from the chemical changes of the sample due to dissociation. To our knowledge, a clear-cut experimental observation of these effects has not been reported.

Using all the spectra for which the peak position can be clearly identified, we find that the relationship between ω_{BP} and the energy of the inherent structure is linear (Fig. 3a). The energy at which ω_{BP} becomes zero, e_c , is found from a linear fit as $e_c = 1.74\epsilon$ (ϵ is the energy scale). As for the height of the peak, the points up to $T = 0.83T_{MC}$ are compatible with a power-law divergence of the form shown in equation (2) (Fig. 3b). Fixing e_c at the value 1.74ϵ arising from the linear fit of ω_{BP} versus e_{IS} , a power-law fit yields an exponent $\beta = 0.040 \pm 0.15$; if the exponent is fixed at $\beta = 1/2$, then the critical value is obtained as $e_c = 1.752 \pm 0.002\epsilon$. Thus the numerical data are compatible with the theoretically predicted scaling, although we have not been able to work close to the critical point, and thus cannot attain great accuracy in the critical exponents or the critical point.

The present discussion applies to experiments as long as the system is in the regime where the inverse frequency is much larger than the structural relaxation time, when the harmonic approximation is applicable³⁰. Thus, although the boson peak observed in liquids can be understood by the present considerations, one should not expect to actually arrive close to the critical point in equilibrium measurements. Rather, one should generate a glass by hyperquenching (see, for example, <http://www.df.unipi.it/workshop/oral/OL6.pdf>): for such a system, we predict that the boson peak measured at low temperature should obey the scaling of equation (2), the control parameter being the fictive temperature (that is, the temperature at which the value of e_{IS} actually reached would be the equilibrium value). In our instantaneous quench, the fictive temperature is simply the temperature at which the hyperquench starts; but in experiments, the cooling rate will also have an effect. In this way, the problem of swamping by the quasielastic peak should also be alleviated.

The significance of the value $e_c = 1.74(1)\epsilon$ is twofold. First, it is the value of e_{IS} that corresponds, in equilibrium, to $T = T_{MC}$, the mode-coupling critical temperature (which we have independently determined as in ref. 22, using a standard Metropolis Monte Carlo without swap moves). Second, it coincides within errors with the threshold energy below which the ratio of the number of saddle points to the number of minima vanishes exponentially with the number of particles⁷ (our numerical accuracy on e_c is similar to that of ref. 7). Thus we find a link between the onset of slow dynamics, the appearance of phonon-like excitations, and a geometrical phase transition of the PEL in fragile glasses. At variance with previous PEL studies, the clear link between the PEL geometrical transition and high-frequency dynamics allows it to be studied experimentally by following the evolution of the boson peak.

We have studied numerically the VDOS of a fragile glass-former undergoing its mode-coupling transition. We have found that in the ensemble of the stationary points near the equilibrium configurations (GISs) there is a transition from a mechanically unstable phase where saddles are present to a stable phase where phonon-like

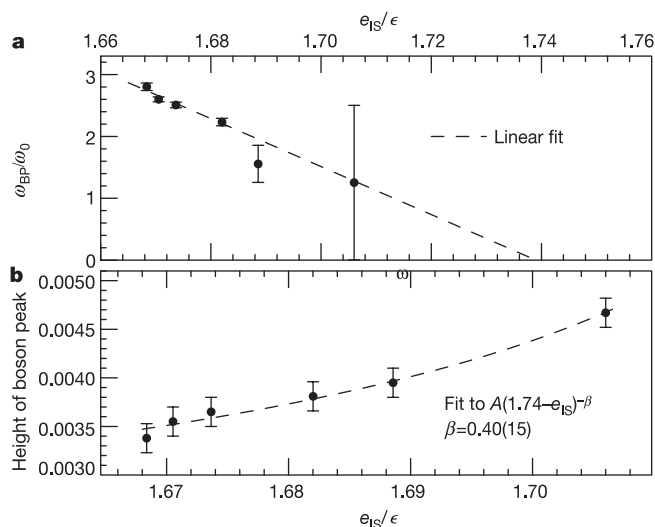


Figure 3 Scaling of the position and height of the boson peak near the liquid-phonon transition. (Energies and frequencies are in units of ϵ and ω_0 , respectively.) **a**, The position of the boson peak, ω_{BP} , is linear in the control parameter, in this case the energy of the inherent structures e_{IS} . Extrapolating with a linear fit, ω_{BP} goes to 0 at $e_{IS} = e_c = 1.74(1)\epsilon$. **b**, The height of the boson peak diverges as a power law (the leftmost point has been left out of the fit). The height and position of the boson peak were obtained by fitting a parabola to the peak of $g(\omega)/\omega^2$ at $T/T_{MC} = 0.49, 0.54, 0.61, 0.69, 0.78$ and 0.83 . The corresponding e_{IS}/ϵ are $1.668, 1.671, 1.674, 1.682, 1.689$ and 1.707 .

excitations appear. The scaling laws predicted by ERMT are compatible with the numerical results that we have obtained for the soft-sphere model. This approach is equivalent to other microscopic descriptions of the glass transition based on the geometry of the PEL^{5–7,31}. But this point of view emphasizes quantities accessible to experiment, like the VDOS, and proposes an interpretation of a universal feature of glasses, the boson peak. This peak appears in the phonon phase, and it is a signature of a cross-over from a phonon-dominated spectrum with a Debye ω^2 scaling to an ω^γ , $\gamma \approx 1.5$ spectrum, resulting from the hybridization of acoustic modes with high-energy modes that soften upon approaching the saddle-phonon transition⁹. The ERMT predictions (equations (1) and (2)) could be checked experimentally in hyperquenching experiments, where the boson peak should strongly depend on the fictive temperature. We expect that the ‘saddle-phonon transition’ point of view will be able to bridge the realms of experiment and numerical studies of the PEL, allowing the testing of many geometrical ideas, and the use of insights derived from PEL in the detailed analysis of the experimental glass transition. □

Received 10 July 2002; accepted 30 January 2003; doi:10.1038/nature01475.

- Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
- DeBenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **410**, 259–267 (2001).
- Sette, F., Krisch, M. H., Masciovecchio, C., Ruocco, G. & Monaco, G. Dynamics of glasses and glass-forming liquids studied by inelastic X-ray scattering. *Science* **280**, 1550–1555 (1998).
- Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **267**, 1935–1939 (1995).
- Angelani, L., Di Leonardo, R., Ruocco, G., Scala, A. & Sciortino, F. Saddles in the energy landscape probed by supercooled liquids. *Phys. Rev. Lett.* **85**, 5356–5359 (2000).
- Broderix, K., Bhattacharya, K. K., Cavagna, A., Zippelius, A. & Giardina, I. Energy landscape of a Lennard-Jones liquid: Statistics of stationary points. *Phys. Rev. Lett.* **85**, 5360–5363 (2000).
- Grigera, T. S., Cavagna, A., Giardina, I. & Parisi, G. Geometric approach to the dynamic glass transition. *Phys. Rev. Lett.* **88**, 055502 (2002).
- Mézard, M., Parisi, G. & Zee, A. Spectra of euclidean random matrices. *Nucl. Phys. B* **559**, 689–701 (1999).
- Grigera, T. S., Martín-Mayor, V., Parisi, G. & Verrocchio, P. Vibrations in glasses and Euclidean random matrix theory. *J. Phys. Condens. Matter* **14**, 2167–2179 (2002).
- Phillips, W. A., Buchenau, U., Nücher, N., Dianoux, A.-J. & Petry, W. Dynamics of glassy and liquid selenium. *Phys. Rev. Lett.* **63**, 2381–2384 (1989).
- Ruocco, G. *et al.* Relaxation processes in harmonic glasses? *Phys. Rev. Lett.* **84**, 5788–5791 (2000).
- Pilla, O. *et al.* Transverse acoustic nature of the excess of vibrational states in vitreous silica. Preprint cond-mat/0209519 at (<http://xxx.lanl.gov>) (2002).
- Masciovecchio, C. *et al.* Observation of large-momentum phononlike modes in glasses. *Phys. Rev. Lett.* **76**, 3356–3359 (1996).
- Matic, A. *et al.* Contrasting behaviour of acoustic modes in network and non-network glasses. *Europhys. Lett.* **54**, 77–83 (2001).
- Masciovecchio, C. *et al.* High-frequency propagating modes in vitreous silica at 295 K. *Phys. Rev. B* **55**, 8049–8051 (1997).
- Benassi, P. *et al.* Evidence of high frequency propagating modes in vitreous silica. *Phys. Rev. Lett.* **77**, 3835–3838 (1996).
- Fioretto, D. *et al.* High-frequency dynamics of glass-forming polybutadiene. *Phys. Rev. E* **59**, 4470–4475 (1999).
- Pilla, O. *et al.* Nature of the short wavelength excitations in vitreous silica: An X-ray Brillouin scattering study. *Phys. Rev. Lett.* **85**, 2136–2139 (2000).
- Martin-Mayor, V., Mézard, M., Parisi, G. & Verrocchio, P. The dynamical structure factor in topologically disordered systems. *J. Chem. Phys.* **114**, 8068–8081 (2001).
- Grigera, T. S., Martín-Mayor, V., Parisi, G. & Verrocchio, P. Vibrational spectrum of topologically disordered systems. *Phys. Rev. Lett.* **87**, 085502 (2001).
- Götze, W. & Sjörgen, L. Relaxation processes in supercooled liquids. *Rep. Prog. Phys.* **55**, 241–376 (1992).
- Kob, W. & Andersen, H. C. Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture I: The van Hove correlation function. *Phys. Rev. E* **51**, 4626–6241 (1995).
- Götze, W. & Mayr, M. Evolution of vibrational excitations in glassy systems. *Phys. Rev. E* **61**, 587–606 (2000).
- Cavagna, A., Giardina, I. & Parisi, G. Role of saddles in mean-field dynamics above the glass transition. *J. Phys. A* **34**, 5317–5326 (2001).
- Bernu, B., Hansen, J.-P., Hiwataru, Y. & Pastore, G. Soft-sphere model for the glass transition in binary alloys: Pair structure and self-diffusion. *Phys. Rev. A* **36**, 4891–4903 (1987).
- Grigera, T. S. & Parisi, G. Fast Monte Carlo algorithm for supercooled soft spheres. *Phys. Rev. E* **63**, 045102 (2001).
- Kob, W., Sciortino, F. & Tartaglia, P. Aging as dynamics in configuration space. *Europhys. Lett.* **49**, 590–596 (2000).
- Engberg, D. *et al.* Origin of the boson peak in a network glass B₂O₃. *Phys. Rev. B* **59**, 4053–4057 (1999).
- Mamedov, S., Kisliuk, A. & Quitmann, D. Effect of preparation conditions on the low frequency Raman spectrum of glassy As₂S₃. *J. Mater. Sci.* **33**, 41–43 (1998).
- Horbach, J., Kob, W. & Binder, K. The specific heat of amorphous silica within the harmonic approximation. *J. Phys. Chem. B* **103**, 4104–4108 (1999).
- La Nave, E., Stanley, H. E. & Sciortino, F. Configuration space connectivity across the fragile-to-strong transition in silica. *Phys. Rev. Lett.* **88**, 035501 (2002).

Acknowledgements We thank O. Pilla, G. Ruocco and G. Viliani for discussions. We are grateful to the RTN3 collaboration for CPU time in their cluster. V.M.M. was supported in part by the European Commission and the Spanish OCYT.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.M.-M. (e-mail: victor.martin@roma1.infn.it).

Detection of human influence on sea-level pressure

Nathan P. Gillett*, Francis W. Zwiers†, Andrew J. Weaver* & Peter A. Stott‡

* School of Earth and Ocean Sciences, University of Victoria, PO Box 3055, Victoria, British Columbia, V8W 3P6, Canada

† Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada, PO Box 1700, STN CSC, Victoria, British Columbia, V8W 2Y2, Canada

‡ Hadley Centre for Climate Prediction and Research, Met Office, Bracknell, Berkshire RG12 2SY, UK

Greenhouse gases and tropospheric sulphate aerosols—the main human influences on climate—have been shown to have had a detectable effect on surface air temperature^{1–3}, the temperature of the free troposphere and stratosphere^{2,4} and ocean temperature^{5,6}. Nevertheless, the question remains as to whether human influence is detectable in any variable other than temperature. Here we detect an influence of anthropogenic greenhouse gases and sulphate aerosols in observations of winter sea-level pressure (December to February), using combined simulations from four climate models. We find increases in sea-level pressure over the subtropical North Atlantic Ocean, southern Europe and North Africa, and decreases in the polar regions and the North Pacific Ocean, in response to human influence. Our analysis also indicates that the climate models substantially underestimate the magnitude of the sea-level pressure response. This discrepancy suggests that the upward trend in the North Atlantic Oscillation index⁷ (corresponding to strengthened westerlies in the North Atlantic region), as simulated in a number of global warming scenarios^{8–10}, may be too small, leading to an underestimation of the impacts of anthropogenic climate change on European climate.

We use gridded observations of decadal mean December–February sea-level pressure (1948–1998) taken from three sources: a version of the HadSLP data set derived using only surface observations¹¹, the National Centers for Environmental Prediction (NCEP) reanalysis¹², and an updated version of the Trenberth data set¹³. Intercomparison of these data sets indicates that agreement is generally good over those regions covered by observations, and that the Trenberth data and NCEP reanalysis differ mainly over Greenland and the Himalayas, probably owing to differences in the reduction of surface pressure to sea level. In regions where there are few surface observations, such as the Antarctic, the NCEP reanalysis is likely to be less reliable than in the better-sampled regions covered by all three data sets. We use ensembles of integrations with historical greenhouse gas and sulphate aerosol forcing from four coupled ocean–atmosphere climate models: the first and second Canadian Centre for Climate Modelling and Analysis coupled models, CGCM1 and CGCM2, and the second and third Hadley Centre coupled models, HadCM2 and HadCM3.

Figure 1 shows the trend in sea-level pressure over the 1948–1998 period in the NCEP reanalysis (Fig. 1a) and the multi-model mean

response (a mean over all available ensemble members) (Fig. 1b). The most notable features of the observed trend are decreases in sea-level pressure over the Arctic, Antarctic and North Pacific, and an increase over the subtropical North Atlantic, southern Europe and North Africa. These features were found to be significant compared to simulated internal variability, and are reproduced in response to greenhouse gas and sulphate aerosol forcing in all four models, and thus also in the multi-model mean (Fig. 1b). The amplitude of the simulated trends is, however, much less than that observed. Some authors have suggested that such a discrepancy might result from deficiencies in the modelled response to sea surface temperature variations¹⁴, or a lack of stratospheric resolution⁸, but such suggestions are controversial, and this issue remains open to debate¹⁵. The difference in simulated and observed trends over the Antarctic is partly attributable to stratospheric ozone depletion: simulated sea-level pressure changes in integrations of HadCM3 that also included stratospheric ozone changes were larger over the Antarctic, but agreed no better with observations over the Northern Hemisphere. The inclusion of solar and volcanic changes also did not resolve this discrepancy.

Until now, model simulations and observations of sea-level pressure changes have only been compared using indices such as the North Atlantic Oscillation index¹⁰. Here we apply an optimal detection approach to gridded sea-level pressure fields, thereby using more spatio-temporal information to assess whether the observed changes are likely to have been due to natural variability, and if not, whether they are consistent with those simulated in response to anthropogenic forcing.

The optimal detection approach³ assumes that the observations (y) may be represented as the linear sum of the scaled simulated

response to greenhouse gases and sulphate aerosol (x), and natural variability (u):

$$y = \beta x + u$$

We apply signal-to-noise optimization, and account for uncertainty in the modelled response pattern by estimating β , the regression coefficient, with a total least-squares fit³. Owing to the limited length of control integration available, we reduce the number of degrees of freedom of our detection vector by regridding to a coarse grid ($60^\circ \times 25^\circ$), and truncating onto the first ten empirical orthogonal functions (EOFs) of control variability³.

We check whether this model provides a plausible explanation of the observations by testing whether the residual, u , is consistent with control variability at the 5% level¹⁶. The uncertainty in β is then assessed from control variability, and the forcing response pattern is detected if β is found to be inconsistent with zero. Averaging results from multiple models has been found to improve estimates of the long-term mean¹⁷ and seasonal forecasts¹⁸ of atmospheric variables, as well as estimates of the surface temperature response to anthropogenic forcing¹⁹. Thus, in order to increase the signal-to-noise ratio and reduce the effect of individual model errors, we use data from all four models simultaneously in the analysis, taking a mean of the simulated responses, and using concatenated control integrations from all four models to estimate internal variability¹⁹.

Figure 2 shows the regression coefficient, β , estimated using this multi-model mean response pattern and each of the three observational data sets. It is significantly greater than zero in each case, indicating that an anthropogenic response is detected. The residual, u , was also found to be consistent with control variability in each case. However, based on 1948–1998 data, the associated 5–95% uncertainty ranges on β do not include unity in any case. This suggests that the models significantly underestimate the amplitude of the sea-level pressure response to greenhouse gas and sulphate aerosol increases, assuming there is no systematic overestimate of the changes in all three observed data sets. This result is consistent with the relative amplitudes of the responses shown in Fig. 1, and was found to be robust to variations in the EOF truncation between 6 and 30, and the use of an ordinary least-squares regression, rather than a total least-squares regression. We also obtained similar results using HadCM2 alone, which has the largest forced ensemble of greenhouse gas and sulphate aerosols, but the response pattern was

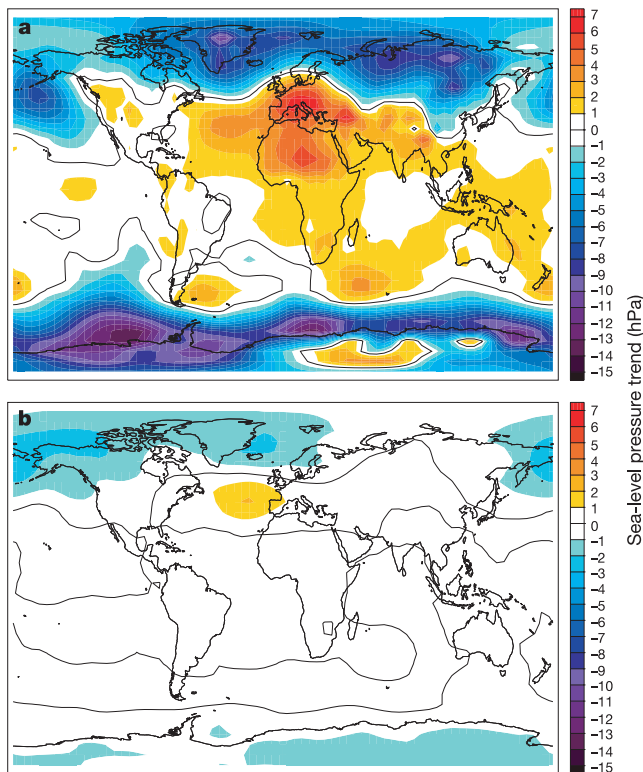


Figure 1 Observed and simulated sea-level pressure trends. The December–February sea-level pressure trends over the period 1948–1998 are shown **a**, for the NCEP reanalysis; and **b**, for the mean of the simulated response to greenhouse gas and sulphate aerosol changes from four climate models (CGCM1, CGCM2, HadCM2 and HadCM3). The pattern of trends is qualitatively similar in each case, but the simulated trends are much smaller in magnitude than those observed.

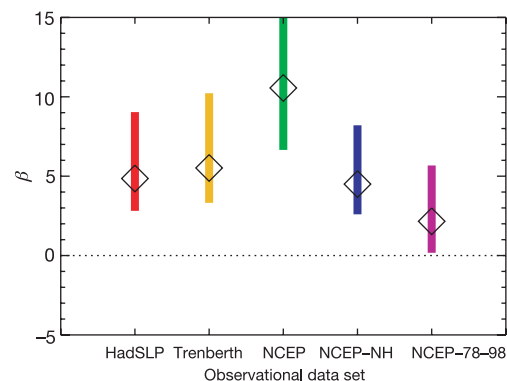


Figure 2 Regression coefficients, β , of observed sea-level pressure changes against changes simulated in response to greenhouse gas and sulphate aerosol increases. The first four bars show regression coefficients derived using a multi-model mean of simulated 1948–1998 sea-level pressure changes, and four observational data sets: HadSLP is a gridded data set of sea-level pressure measurements, defined only where observations are present¹¹; Trenberth is a synthesis of analyses, only defined northward of 20°N ¹³; NCEP is a global reanalysis¹²; and NCEP-NH is the NCEP reanalysis restricted to the Northern Hemisphere. The bar labelled NCEP-78-98 was derived using NCEP reanalysis data from 1978 to 1998 only. The 5–95% uncertainty ranges shown are derived from control variability and the squares represent best estimates.

not detected in every case when we used the other models individually.

Results based on the NCEP reanalysis were more consistent with those derived using the other data sets when only Northern Hemisphere data were used (Fig. 2), probably because of the effects of stratospheric ozone depletion and possible data problems over the Antarctic. To address the issue of possible discontinuities associated with the introduction of satellite data in the NCEP reanalysis around 1979, we repeated the analysis using only the final two decades of NCEP data (1978–1998), and found that the greenhouse gas and sulphate aerosol response was still detectable (Fig. 2).

Overall, we find that anthropogenic greenhouse gases and sulphate aerosols have had a detectable influence on sea-level pressure over the second half of the twentieth century: this represents evidence of human influence on climate independent of measurements of temperature change. We find that the observed pattern of December–February sea-level pressure trends is similar to that simulated in four climate models between 1948 and 1998, with decreasing sea-level pressure over the poles and North Pacific, and an increase over the subtropical North Atlantic. However, our results also indicate that these climate models may be substantially underpredicting this sea-level pressure response.

Circulation trends have had important regional impacts on climate: for example, the trend in the North Atlantic Oscillation has been associated with approximately 50% of the observed Eurasian winter warming over the past thirty years, over 60% of the rainfall increase in Scotland, and over 60% of the rainfall decrease in Spain over the same period⁷. It has also been linked to large changes in extreme events, for example over 70% of the decrease in extreme cold events in France²⁰. On the basis of the results presented here, we conclude that such effects are probably underestimated in model simulations of both current and future climate change, although this does not invalidate the attribution of surface temperature trends to human influence²¹. However, if we are to make realistic predictions of regional climate change, it is important to reconcile this apparent discrepancy between models and observations. □

Received 10 September 2002; accepted 7 February 2003; doi:10.1038/nature01487.

1. Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. & Mitchell, J. F. B. Causes of twentieth-century temperature change near the Earth's surface. *Nature* **399**, 569–572 (1999).
2. Mitchell, J. F. B. *et al.* *Climate Change 2001. The Scientific Basis* Ch. 12 (Cambridge Univ. Press, Cambridge, UK, 2001).
3. Allen, M. R. *et al.* Quantifying anthropogenic influence on recent near-surface temperature change. *Serv. Geophys.* (in the press).
4. Tett, S. F. B., Mitchell, J. F. B., Parker, D. E. & Allen, M. R. Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science* **274**, 1170–1173 (1996).
5. Barnett, T. P., Pierce, D. W. & Schnur, R. Detection of anthropogenic climate change in the world's oceans. *Science* **292**, 270–274 (2001).
6. Reichert, B. K. *et al.* Global ocean warming tied to anthropogenic forcing. *Geophys. Res. Lett.* DOI:10.1029/2001GL013954 (2002).
7. Thompson, D. W. J., Wallace, J. M. & Hegerl, G. C. Annular modes in the extratropical circulation. Part II: Trends. *J. Clim.* **13**, 1018–1036 (2000).
8. Shindell, D. T., Miller, R. L., Schmidt, G. A. & Pandolfo, L. Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature* **399**, 452–455 (1999).
9. Fyfe, J. C., Boer, G. J. & Flato, G. M. The Arctic and Antarctic oscillations and their projected changes under global warming. *Geophys. Res. Lett.* **26**, 1601–1604 (1999).
10. Gillett, N. P. *et al.* How linear is the Arctic Oscillation response to greenhouse gases? *J. Geophys. Res.* DOI:10.1029/2001JD000589 (2002).
11. Basnett, T. A. & Parker, D. E. Development of the global mean sea-level pressure data set GMSLP2. *Clim. Res. Tech. Note 79* (Hadley Centre, Met Office, Bracknell, UK, 1997).
12. Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
13. Trenberth, K. E. & Paolino, D. A. The Northern Hemisphere sea-level pressure data set: Trends, errors and discontinuities. *Mon. Weath. Rev.* **108**, 855–872 (1980).
14. Osborn, T. J. *et al.* Evaluation of the North Atlantic Oscillation as simulated by a coupled climate model. *Clim. Dyn.* **15**, 685–702 (1999).
15. Gillett, N. P., Graf, H. F. & Osborn, T. J. in *The North Atlantic Oscillation: Climatic Significance and Environmental Impact* Ch. 9 (American Geophysical Union, Washington DC, 2003).
16. Allen, M. R. & Tett, S. F. B. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* **15**, 419–434 (1999).
17. Lambert, S. J. & Boer, G. J. CMIP1 evaluation and intercomparison of coupled climate models. *Clim. Dyn.* **17**, 83–106 (2001).

18. Krishnamurti, T. N. *et al.* Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* **285**, 1548–1550 (1999).
19. Gillett, N. P. *et al.* Detecting anthropogenic influence with a multi-model ensemble. *Geophys. Res. Lett.* DOI:10.1029/2002GL015836 (2002).
20. Thompson, D. W. J. & Wallace, J. M. Regional climate impacts of the Northern Hemisphere annular mode. *Science* **293**, 85–89 (2001).
21. Gillett, N. P. *et al.* Implications of changes in the Northern Hemisphere circulation for the detection of anthropogenic climate change. *Geophys. Res. Lett.* **27**, 993–996 (2000).

Acknowledgements We thank R. Allan and J. Arnott for supplying the HadSLP data set, and M. Allen for allowing us to use his optimal detection code. We thank NSERC and CFCAS for CLIVAR funding. A.J.W. is grateful for support from the Killam Foundation and the Canada Research Chair Program, and P.A.S. is grateful for support from the Department for Environment, Food and Rural Affairs.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.P.G. (e-mail: gillett@uvic.ca).

Enhanced mantle-to-crust rhenium transfer in undegassed arc magmas

Weidong Sun*, Vickie C. Bennett*, Stephen M. Eggins*, Vadim S. Kamenetsky† & Richard J. Arculus‡

* Research School of Earth Sciences, ‡ Department of Geology, The Australian National University, Canberra, ACT 0200, Australia

† School of Earth Sciences and CODES SRC, University of Tasmania, Hobart, Tasmania 7001, Australia

Variations in the ¹⁸⁷Os/¹⁸⁸Os isotopic signature of mantle and mantle-derived rocks have been thought to provide a powerful chemical tracer of deep Earth structure. Many studies have inferred from such data that a long-lived, high-rhenium component exists in the deep mantle (¹⁸⁷Re is the parent isotope decaying to ¹⁸⁷Os, with a half-life of ~42 billion years), and that this reservoir probably consists of subducted oceanic crust^{1–3}. The interpretation of these isotopic signatures is, however, dependent on accurate estimates of rhenium and osmium concentrations in all of the main geochemical reservoirs, and the crust has generally been considered to be a minor contributor to such global budgets. In contrast, we here present observations of high rhenium concentrations and low Yb/Re ratios in arc-type melt inclusions. These results indicate strong enrichment of rhenium in undegassed arc rocks, and consequently the continental crust, which results in a crustal estimate of ~2 p.p.b. rhenium, as compared to previous estimates of 0.4–0.2 p.p.b. (refs 4, 5). Previous determinations of rhenium in arc materials, which were largely measured on subaerially erupted samples, are likely to be in error owing to rhenium loss during degassing. High mantle-to-crust rhenium fluxes, as observed here, require a revaluation of geochemical models based on the ¹⁸⁷Re–¹⁸⁷Os decay system^{1–3}.

The major geochemical reservoirs of mid-ocean-ridge basalts (MORBs), depleted MORB mantle (DMM), continental crust (CC) and primitive mantle (PM) are not complementary in terms of Re contents or Re/lithophile-element ratios. For example, Re and the lithophile element Yb are similarly incompatible in the upper mantle during generation of MORB magma^{1,6}, which is best illustrated by the narrow range of Yb/Re (refs 1, 6; Fig. 1). This suggests that there is no major fractionation between Re and Yb during formation of MORB, and furthermore that the Yb/Re ratio

of the parental material, the DMM, is similar ($3.6 \text{ p.p.m. p.p.b.}^{-1}$)⁶, with both clearly higher than that of the PM ($1.8 \text{ p.p.m. p.p.b.}^{-1}$)⁷ (Fig. 1). The currently accepted estimates of Re abundance in the CC are 0.4 p.p.b. (ref. 4) to 0.2 p.p.b. (ref. 5), which can be translated into average Yb/Re of $5\text{--}10 \text{ (p.p.m. p.p.b.}^{-1})$ using estimates of Yb in the CC of 2 p.p.m. (ref. 8). Thus the major geochemical reservoirs—MORB, DMM and CC—all have Yb/Re values higher than that of the PM, leaving an apparent missing Re component¹. These observations require either that there is an unrecognized major, low-Yb/Re geochemical domain, or that estimates of Re distributions are inaccurate. Numerous studies have demonstrated that at least parts of the upper mantle are characterized by low $^{187}\text{Os}/^{188}\text{Os}$ isotopic compositions relative to PM, reflecting integrated low Re/Os. Dependent on the average $^{187}\text{Os}/^{188}\text{Os}$ of the DMM assumed, various studies propose a mass of recycled MORB equivalent to $2\text{--}22\%$ of the total mantle is stored in the lower mantle^{1–3}. Storage of subducted basaltic crust in the deep mantle can account for the absolute amount of Re needed to balance long-term Re/Os depletions in the DMM, but because recycled MORB would have a Yb/Re equal to, or greater than, fresh MORB (owing to preferential Re loss from the slab during subduction processes⁹), the problem of a complementary low-Yb/Re reservoir remains unresolved.

Alternatively, the estimated Re and Yb/Re characteristics of one or more of the principal reservoirs may be in error. The complicated geochemical behaviour of Re makes it difficult to estimate its abundance in the crust using conventional methods successfully applied to other, generally lithophile elements. Rhenium variously exhibits redox-sensitive, volatile, chalcophile, siderophile and moderately incompatible trace-element behaviour. Rhenium can also be very mobile under surface conditions, similar to Mo and U and other oxyanions^{4,10,11}. The overall abundance in the silicate Earth is very low (0.28 p.p.b. in the PM)⁷, but Re can be enriched to weight per cent levels in some sulphides (for example, molybdenite and pure Re sulphides)^{12,13} and to p.p.m. levels in anoxic sediments^{10,11}. This results in the Re distribution in the crust being highly heterogeneous and strongly biased by the distribution of Re-enriched sulphide, black shale and other anoxic sediments.

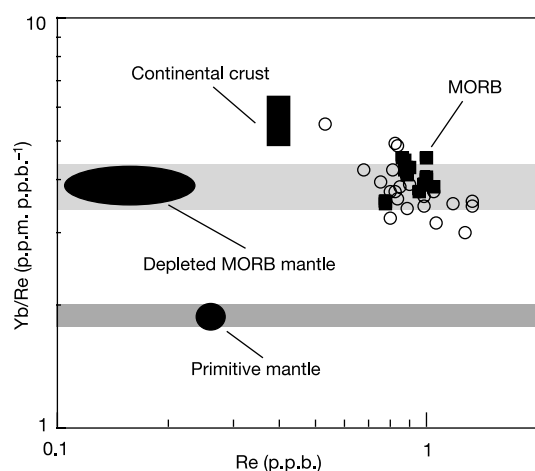


Figure 1 Yb/Re ratio as a function of Re abundance. Data points are from MORB glasses; also shown are estimates for primitive mantle⁷ and continental crust⁴. The constant Yb/Re ratios of MORB indicate similar behaviour of the two elements during magmatic evolution, and provide an estimate of the Yb/Re ratio of their mantle source (depleted MORB mantle)^{1,6}. The major geochemical reservoirs are not complementary, with all having Yb/Re lying above the PM values. This suggests the existence of an additional low-Yb/Re, high-Re source. Data sources: MORB laser ablation ICP-MS data (filled squares), ref. 6; MORB isotope dilution Re data (open circles), ref. 30. Corresponding Yb concentrations for the latter are calculated from Tb abundances (ref. 30) assuming a chondritic Yb/Tb value.

Magmas produced in arcs as the result of subduction processes are generally considered to be the major building blocks of the continental crust, such that analysis of Re in arc materials should provide an alternative, and more direct, estimate of Re transfer from the mantle to the crust. Previously published data from arc volcanic rocks, which are represented by largely subaerially erupted samples, indicate very low Re concentrations ($0.01\text{--}1.67 \text{ p.p.b.}$ with an average of $0.30 \pm 0.02 \text{ p.p.b.}$)^{14–20}. Recent comparisons of subaerially (degassed) and submarine (undegassed) erupted Hawaiian oceanic basalt samples have demonstrated that volatile Re can be lost even from extremely fresh, recently erupted, lavas by magma degassing, and has probably resulted in previous underestimates of Re concentrations in modern plume magmas^{21,22}. This suggests that the apparent low Re in arcs may be a sampling artefact, particularly as the effects of Re loss through magma degassing are likely to be more pronounced in the more volatile-rich arc magmas. That substantial Re loss from arcs does occur is evident both from measurement of high Re in magmatic gases and in the presence of pure Re sublimates in arc systems¹². To address this problem, we determined Re concentrations of primitive subduction-related magmas (Supplementary Information), contained as inclusions in high-magnesium olivine (forsterite content, $\text{Fo} > 85 \text{ mol}\%$) within subaerial arc picritic basalts (Aoba, Vanuatu)²³ or within submarine backarc basaltic rocks with clear arc characteristics (eastern Manus basin, and Valu Fa ridge^{24,25}). The determinations were done *in situ*, using laser ablation inductively coupled plasma-mass spectrometry (ICP-MS); we used an excimer laser coupled to an Agilent 7500 ICP-MS. Analyses were made using a large-diameter spot ($130\text{--}220 \mu\text{m}$, depending on the size of the melt inclusions) and fast repetition rate ($15\text{--}20 \text{ Hz}$)^{6,26}. This approach was used for two reasons; first, to avoid problems associated with Re loss due to degassing (both the high water overpressure for these submarine erupted samples and the enclosing olivine crystals prevent degassing), and second, to avoid the effects of secondary mineral fractionation of Re that might be experienced in more evolved arc magmas²⁰.

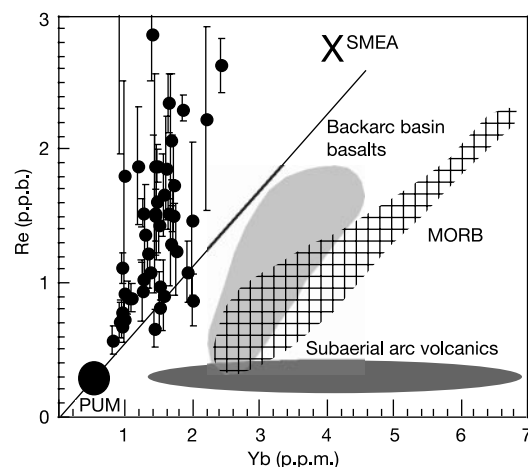


Figure 2 Re abundance versus Yb abundance. Data points are from melt inclusions hosted by olivines from arc picrites and backarc basalts of clear arc characteristics; these data were obtained by laser ablation ICP-MS. Also shown are fields for MORBs, primitive upper mantle (PUM), backarc basin basalts, the average of previously published results for subaerially erupted arc volcanics, and the average of submarine erupted, evolved arc lavas ($\text{MgO} < 3.5 \text{ wt}\%$) (SMEA). The melt inclusions (filled circles with error bars) are enriched in Re, and provide strong evidence that arcs and ultimately the continental crust are more enriched in Re and with lower Yb/Re ratios than previously estimated. Data sources: subaerially erupted arc volcanics, refs 14–20; backarc basin basalts, ref. 6; PUM, ref. 7; MORB, refs 6, 30; SMEA, W.S., R.J.A., V.C.B., S.M.E. and R. A. Binns, manuscript in preparation).

All the Re concentrations that we measured in the melt inclusions are substantially higher than previously observed in arc samples^{14–20}, and range from 0.65 to 6.0 p.p.b., with less variation in Yb (0.97–2.86 p.p.m.) (Fig. 2; Supplementary Information). The Yb/Re ratio ranges from 0.18 to 2.33 p.p.m. p.p.b.^{–1}, with an average of 1.00 ± 0.07 p.p.m. p.p.b.^{–1}, and which is significantly lower than that of the PM (1.8 p.p.m. p.p.b.^{–1}; ref. 7) or previous crustal estimates (5–10 p.p.m. p.p.b.^{–1}) (Figs 1 and 3). These results from three well-documented southern Pacific localities demonstrate regionally high rates of Re transfer from the mantle, via arc magmatism. We anticipate that additional work will confirm that the pattern of undegassed arc materials having higher Re is consistent throughout other arc systems. The large variation of Yb/Re (Fig. 3) in the inclusions may be due to different degrees of Re enrichment through addition of subduction-derived fluids and/or early Re loss. Overall, the melt inclusions are expected to be well preserved from degassing and sulphide segregation, but some pre-entrapment Re loss cannot be excluded. Therefore, the Yb/Re ratio of the parent magma of the melt inclusions might be even lower than the weighted average (1.00 ± 0.07 p.p.m. p.p.b.^{–1}).

The post-Archaeon CC is primarily a product of convergent margin magmatism with minor contributions from plume-related basalts^{8,27}, both of which have low Yb/Re ratios^{6,21}. Assuming that the Yb/Re ratio of the CC is similar to the average of the arc melt inclusions (1.00 ± 0.07 p.p.m. p.p.b.^{–1}), and using an average crustal Yb abundance of 2 p.p.m. (ref. 8), we obtain a Re abundance of 2 p.p.b.—this is a factor of 5–10 greater than previous estimates. Similar concentrations (average 1.6 ± 0.14 p.p.b.) are measured directly for the 40 melt inclusions (Supplementary Information). How closely can we relate the Re compositions in the primitive arc melt inclusions to average crustal compositions that are chemically more similar to the evolved parts of arc systems? Evidence that high Re is characteristic of the entire undegassed arc system is provided by analyses of Re in evolved submarine erupted glasses (MgO, 0.33–3.50 wt%) from the Manus basin, which yield an average Re content of 3.1 ± 1.5 p.p.b. ($n = 9$; W.S., R.J.A., V.C.B., S.M.E. and R. A. Binns, manuscript in preparation).

The applicability of the new Re estimates from arcs for the total preserved CC, and their importance for global Re–Os budgets, depend largely on the amount of Re lost from arcs by degassing and on the ultimate fate of this lost Re. A large proportion of the degassed Re will probably have been transported to the ocean, and then ultimately incorporated into suboxic and anoxic sediments^{10,11,28}. Loss of Re during arc volcanism may explain the apparent imbalance in oceanic sedimentary sinks of Re, which in some cases are estimated to contain $\geq 100\%$ of the riverine inputs of

Re (refs 10, 11). If either relatively small amounts of Re are lost from arcs by degassing, or if the sedimentary Re sinks are ultimately incorporated into the continents at convergent margins, then Re may be expected to have a long crustal residence time—of the order of the mean continental age of ~ 2 Gyr. This has important consequences for global Re–Os systematics. First, it has been well demonstrated that some portions of the DMM must be Re-depleted with respect to PM compositions (see for example, refs 1, 6), although there is still no consensus as to the average degree of depletion. Depending on the amount of Re depletion assumed for the DMM, a high-Re continental crust can account for all (or at least a significant portion) of the missing Re, decreasing or eliminating the need for storage of recycled MORB to provide the conjugate Re-enriched reservoir. Second, a higher Re crustal abundance presents challenges to our current understanding of crustal Os isotope compositions, in that concomitant higher Os abundances would be needed to balance the estimated average ¹⁸⁷Os/¹⁸⁸Os ratio of the CC^{4,5}. However, Os is similar to Re in being extremely heterogeneously distributed within the crust and being based on relatively few data, leaving large uncertainties associated with current average crustal estimates.

Alternatively, substantial amounts of Re may be lost by degassing, with the Re hosts (anoxic sediments) quickly subducted, resulting in a short (< 200 Myr) crustal residence time. In this scheme, the average Re content of the preserved continental crust will be between 0.2 (current estimate) and 2 p.p.b., and the role of the preserved crust as a conjugate reservoir to balance mantle depletions will be less significant. Subducted oceanic crust comprises MORB, oceanic lithospheric mantle and associated sediments²⁹. If anoxic sediments are a primary sink for Re lost from arcs, then subduction and deep-mantle storage of anoxic sediments, with their high Re and low Yb/Re, may provide the missing component in global Re–Os budgets (Fig. 1). In addition, subducted high-Re sediments may create local areas of enriched Re/Os and ultimately high Os isotopic heterogeneity in the mantle.

The compositions of undegassed arc magmas provide a strong basis for upwards revision of Re estimates for the continental crust, suggesting concentrations as high as 2 p.p.b. Regardless of the degree of loss during arc eruptions and the crustal residence time of Re, the new data presented here demonstrate that arc processes can transport large amounts of Re from the mantle to the crust. Much of the Re in arcs may ultimately be preserved within the continents, with the continental crust a more significant repository for Re than previously recognized, and with a distinctively low Yb/Re ratio. The present observations reveal complexities in the Re cycle with direct consequences for global Re–Os mass balances. □

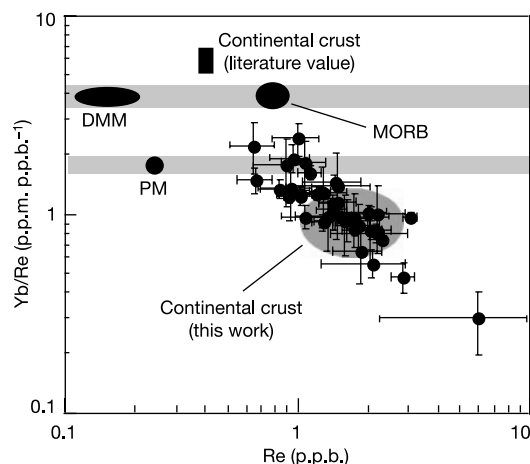


Figure 3 Revised estimates of Re and Yb/Re in the continental crust (this work). These data are compared with previous estimates for the continental crust, and with PM, DMM and MORB values. Data sources as in Fig. 1. Melt inclusions are shown as filled circles with error bars.

Received 22 October 2002; accepted 30 January 2003; doi:10.1038/nature01482.

- Hauri, E. H. & Hart, S. R. Rhenium abundances and systematics in oceanic basalts. *Chem. Geol.* **139**, 185–205 (1997).
- Walker, R. J., Prichard, H. M., Ishiwatari, A. & Pimentel, M. The osmium isotopic composition of convecting upper mantle deduced from ophiolite chromites. *Geochim. Cosmochim. Acta* **66**, 329–345 (2002).
- Bennett, V. C., Nutman, A. P. & Esat, T. M. Constraints on mantle evolution from $^{187}\text{Os}/^{188}\text{Os}$ isotope compositions of Archean ultramafic rocks from southern West Greenland (3.8 Ga) and Western Australia (3.46 Ga). *Geochim. Cosmochim. Acta* **66**, 2615–2630 (2002).
- Esser, B. K. & Turekian, K. K. The osmium isotopic composition of the continental crust. *Geochim. Cosmochim. Acta* **57**, 3093–3104 (1993).
- Peucker-Ehrenbrink, B. & Jahn, B. Rhenium-osmium isotope systematics and platinum group element concentrations; loess and the upper continental crust. *Geochim. Geophys. Geosyst.* **2**, 2001GC000172 (2001).
- Sun, W. D., Bennett, V. C., Eggins, S. M., Arculus, R. J. & Perfit, M. R. Rhenium systematics in submarine MORB and back-arc basin glasses: Laser ablation ICP-MS results. *Chem. Geol.* (in the press).
- McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust; a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
- Becker, H. Re-Os fractionation in eclogites and blueschists and the implications for recycling of oceanic crust into the mantle. *Earth Planet. Sci. Lett.* **177**, 287–300 (2000).
- Morford, J. L. & Emerson, S. The geochemistry of redox sensitive trace metals in sediments. *Geochim. Cosmochim. Acta* **63**, 1735–1750 (1999).
- Jaffe, L. A., Peucker-Ehrenbrink, B. & Petsch, S. T. Mobility of rhenium, platinum group elements and organic carbon during black shale weathering. *Earth Planet. Sci. Lett.* **198**, 339–353 (2002).
- Korzhinsky, M. A., Tkachenko, S. I., Shmulovich, K. I., Taran, Y. A. & Steinberg, G. S. Discovery of a pure rhenium mineral at Kudriav Volcano. *Nature* **369**, 51–52 (1994).
- Ishihara, S. Rhenium contents of molybdenites in granitoid-series rocks in Japan. *Econ. Geol. Bull. Soc. Econ. Geol.* **83**, 1047–1051 (1988).
- Alves, S., Schiano, P., Capmas, F. & Allègre, C. J. Osmium isotope binary mixing arrays in arc volcanism. *Earth Planet. Sci. Lett.* **198**, 355–369 (2002).
- Alves, S., Schiano, P. & Allègre, C. J. Rhenium-osmium isotopic investigation of Java subduction zone lavas. *Earth Planet. Sci. Lett.* **168**, 65–77 (1999).
- Woodland, S. J., Pearson, D. G. & Thirlwall, M. F. A platinum group element and Re-Os isotope investigation of siderophile element recycling in subduction zones: Comparison of Grenada, Lesser Antilles Arc and the Izu-Bonin Arc. *J. Petrol.* **42**, 171–198 (2002).
- Lassiter, J. C. & Luhr, J. F. Osmium abundance and isotope variations in mafic Mexican volcanic rocks: Evidence for crustal contamination and constraints on the geochemical behavior of osmium during partial melting and fractional crystallization. *Geochim. Geophys. Geosyst.* **2**, 2000GC000116 (2001).
- Chesley, J. T., Ruiz, J., Righter, K., Ferrari, L. & Gomez-Tuena, A. Source contamination versus assimilation: An example from the Trans-Mexican Volcanic Arc. *Earth Planet. Sci. Lett.* **195**, 211–221 (2002).
- Borg, L. E., Brandon, A. D., Clynne, M. A. & Walker, R. J. Re-Os isotopic systematics of primitive lavas from the Lassen region of the Cascade Arc, California. *Earth Planet. Sci. Lett.* **177**, 301–317 (2000).
- Righter, K., Chesley, J. T. & Ruiz, J. Genesis of primitive, arc-type basalt: Constraints from Re, Os, and Cl on the depth of melting and role of fluids. *Geology* **30**, 619–622 (2002).
- Bennett, V. C., Norman, M. D. & Garcia, M. O. Rhenium and platinum group element abundances correlated with mantle source components in Hawaiian picrites; sulphides in the plume. *Earth Planet. Sci. Lett.* **183**, 513–526 (2000).
- Lassiter, J. C. Evidence for Re loss during magma degassing: Implications for Re concentration variations in OIB and MORB. *Eos* **83**(47) (Fall Meet. Suppl.), 1447 (2002).
- Eggins, S. M. Origin and differentiation of picritic arc magmas, Ambae (Aoba), Vanuatu. *Contrib. Mineral. Petrol.* **114**, 79–100 (1993).
- Kamenetsky, V. S., Crawford, A. J., Eggins, S. M. & Muehe, R. Phenocryst and melt inclusion chemistry of near-axis seamounts, Valu Fa Ridge, Lau Basin; insight into mantle wedge melting and the addition of subduction components. *Earth Planet. Sci. Lett.* **151**, 205–223 (1997).
- Kamenetsky, V. S. *et al.* Parental basaltic melts and fluids in eastern Manus backarc basin; implications for hydrothermal mineralisation. *Earth Planet. Sci. Lett.* **184**, 685–702 (2001).
- Sylvester, P. J. & Eggins, S. M. Analysis of Re, Au, Pd, Pt and Rh in NIST glass certified reference materials and natural basalt glasses by laser ablation ICP-MS. *Geostand. Newsl.* **21**, 215–229 (1997).
- Barth, M. G., McDonough, W. F. & Rudnick, R. L. Tracking the budget of Nb and Ta in the continental crust. *Chem. Geol.* **165**, 197–213 (2000).
- Peucker-Ehrenbrink, B. & Hannigan, R. E. Effects of black shale weathering on the mobility of rhenium and platinum group elements. *Geology* **28**, 475–478 (2000).
- Hofmann, A. W. Mantle geochemistry; the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Schiano, P., Birc, J. L. & Allègre, C. J. Osmium-strontium-neodymium-lead isotopic covariations in mid-ocean ridge basalt glasses and the heterogeneity of the upper mantle. *Earth Planet. Sci. Lett.* **150**, 363–379 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank M. Kamenetsky, C. Allen and J. M. G. Shelley for help with sample preparation and laser ablation ICP-MS analyses. We also thank S.-s. Sun for discussions that initiated this work, and D. G. Pearson for comments and suggestions that improved the manuscript. W.S. was supported by an Australian international graduate student scholarship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.C.B. (e-mail: vickie.bennett@anu.edu.au).

Genetics of gene expression surveyed in maize, mouse and man

Eric E. Schadt^{*†}, Stephanie A. Monks^{*†‡}, Thomas A. Drake[§], Aldons J. Lusis^{||}, Nam Che^{||}, Veronica Colina^{||}, Thomas G. Ruff[¶], Stephen B. Milligan^{*}, John R. Lamb^{*}, Guy Cavet^{*}, Peter S. Linsley^{*}, Mao Mao^{*}, Roland B. Stoughton^{*} & Stephen H. Friend^{*#}

^{*} Rosetta Inpharmatics, LLC, 12040 115th Avenue N.E., Kirkland, Washington 98034, USA

[‡] Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA

[§] Department of Pathology and Laboratory Medicine, and ^{||} Department of Microbiology, Molecular Genetics and Immunology, Department of Medicine, and Department of Human Genetics, UCLA, Los Angeles, California 90095, USA

[¶] Monsanto Company, St Louis, Missouri 63167, USA

[#] Merck Research Laboratories, Merck & Co., Inc., POB 4, WP14-2500, West Point, Pennsylvania 19486, USA

[†] These authors contributed equally to this work

Treating messenger RNA transcript abundances as quantitative traits and mapping gene expression quantitative trait loci for these traits has been pursued in gene-specific ways. Transcript abundances often serve as a surrogate for classical quantitative traits in that the levels of expression are significantly correlated with the classical traits across members of a segregating population. The correlation structure between transcript abundances and classical traits has been used to identify susceptibility loci for complex diseases such as diabetes¹ and allergic asthma². One study recently completed the first comprehensive dissection of transcriptional regulation in budding yeast³, giving a detailed glimpse of a genome-wide survey of the genetics of gene expression. Unlike classical quantitative traits, which often represent gross clinical measurements that may be far removed from the biological processes giving rise to them, the genetic linkages associated with transcript abundance affords a closer look at cellular biochemical processes. Here we describe comprehensive genetic screens of mouse, plant and human transcriptomes by considering gene expression values as quantitative traits. We identify a gene expression pattern strongly associated with obesity in a murine cross, and observe two distinct obesity subtypes. Furthermore, we find that these obesity subtypes are under the control of different loci.

Liver tissues from 111 F₂ mice constructed from two standard inbred strains, C57BL/6J and DBA/2J, were profiled using a mouse gene oligonucleotide microarray. The expression values from these experiments were treated as quantitative traits and carried through a linkage analysis using evenly spaced markers across the autosomal chromosomes. Of the 23,574 genes represented on the microarray, 7,861 were detected as significantly differentially expressed (type I error = 0.05) in the parental strains or in at least 10% of the F₂ mice profiled. Using standard interval mapping techniques, quantitative trait loci (QTL) with log of the odds ratio (LOD) scores greater than 4.3 (P -value < 0.00005) were identified for 2,123 genes, with a maximum LOD score of 80.0 (P -value $\ll 10^{-20}$). On average, gene expression QTL (eQTL) with LOD scores greater than 4.3 explained 25% of the transcription variation of the corresponding genes observed in this F₂ set, with this percentage increasing to nearly 50% for LOD scores greater than 7.0. Applying a simple Bonferroni correction to a 0.05 genome-wide significance level for each trait, given that 7,861 traits were tested, we would expect only 393 false positives at a LOD score threshold of 4.3.

Without filtering based on significant differential expression over the set of mice profiled, we detected 4,339 eQTL over 3,701 genes with LOD scores greater than 4.3, 11,021 genes with at least one

eQTL with a LOD score greater than 3.0, and 17,415 eQTL with LOD scores over 2.0. The number of eQTL with LOD scores exceeding 7.0 (P -value $\leq 10^{-7}$) increased by 50% when genes that were not detected as significantly differentially regulated, as defined above, were considered. This result indicates that although individual tests of hypotheses on the differential regulation of a single gene may not be significant, viewing the behaviour of that gene by genotype over 111 animals provides significantly more information on the biological activity of that gene, an observation also noted for yeast³. Of the 965 genes with LOD scores greater than 7.0, 157 had a maximum fold change of 3.0, indicating a class of genes whose high LOD scores reflect tight transcriptional control (small variance), not large expression differences.

Figure 1a plots the percentage of eQTL at different LOD score thresholds across 920 evenly spaced bins, each 2 cM wide, covering the mouse genome. The number of eQTL in each bin was divided by

the total number of eQTL and plotted. eQTL hotspots are apparent on chromosomes 2, 6, 7, 9, 10, 16 and 17, where for each of these hotspot locations, greater than 1% of the total number of eQTL identified genome-wide localize to a 4-cM window. With 460 4-cM windows over the 19 autosomal chromosomes, the probability that greater than 1% of the eQTL would localize to one such window is less than 1.0×10^{-16} . Furthermore, gene expression is seen to be a complex trait, given that 40% of genes with at least one eQTL with a LOD score greater than 3.0 had more than one eQTL, and close to 4% of such genes had more than three eQTL. It is the clustering of eQTL to specific loci and the relationship between these genes with respect to expression, rather than single eQTL significance levels, that when taken together lead to highly significant and interesting patterns that can be associated with phenotypes related to common diseases (see below).

Of the 18,460 genes that could be mapped to a chromosome

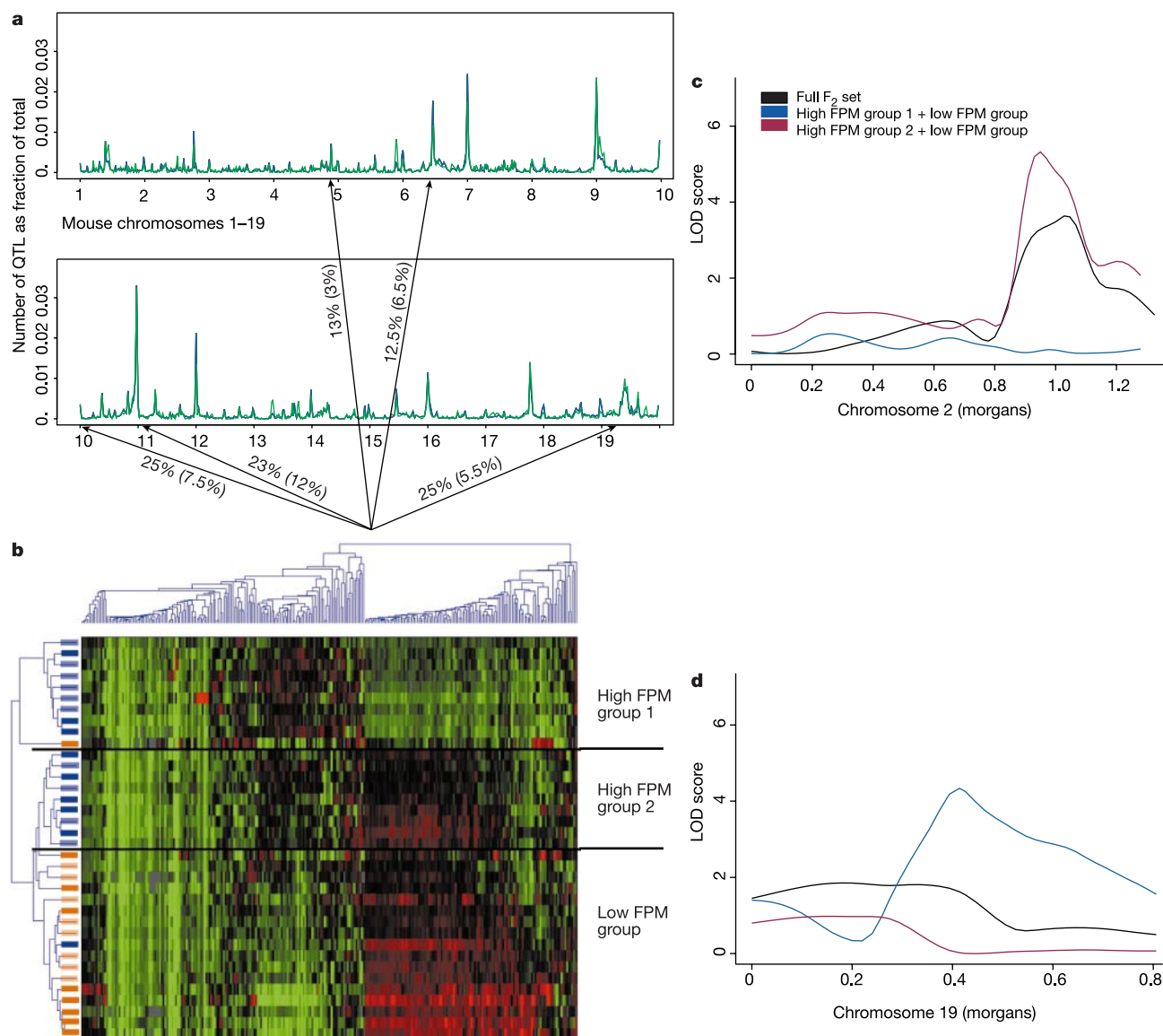


Figure 1 Murine gene expression quantitative trait loci (eQTL) distributions and the molecular basis for fat pad mass (FPM) in a murine F₂ cross. **a**, Percentage of eQTL in 2-cM bins spanning the murine autosomal chromosomes at two LOD score thresholds. **b**, Colour matrix display for hierarchically clustered genes (x axis) and extreme FPM mice (y axis). Dark/light blue bars indicate mice in the upper/lower half of the high FPM group; dark/light orange bars indicate mice in the lower/upper half of the low FPM group.

See text for a definition of the five arrows. **c**, **d**, Chromosome 2 (**c**) and 19 (**d**) log of the odds ratio (LOD) score curves for the FPM trait. QTL analysis was performed on three different sets: (1) black curves, all F₂ mice; (2) blue curves, F₂ mice classified as high FPM group 1 and low FPM group; (3) red curves, F₂ mice classified as high FPM group 2 and low FPM group.

position using the Celera Mouse Genome Database, 3,007 had eQTL with LOD scores greater than 4.3, and 784 had eQTL with LOD scores greater than 7.0. Approximately 34% of the mapped genes with eQTL exceeding 4.3 had a physical location coincident with the eQTL position, whereas 71% of the mapped genes with eQTL exceeding 7.0 had a physical location coincident with its eQTL position. The trend observed here is that eQTL with high LOD scores are *cis*-acting in most cases, whereas moderately significant QTL are *trans*-acting in most cases. This is consistent with our expectation that first-order effects (DNA variations in a gene that affect transcription of the gene itself) are easier to detect than second-order effects (genes acting on the transcription of other genes).

Figure 2 highlights a range of interesting gene-centred polymorphisms known to exist between the DBA and B6 strains (*cis*-acting transcription regulation). The *C5* gene (NM_010406) has a 2-base-pair (bp) deletion in the coding region in DBA mice that leads to rapid decay of the transcript², compared with B6. A LOD of 27.4 centred over the *C5* gene on chromosome 2 is readily detected (black curve). The *Alad* (AK002300) gene is present in two copies in DBA and one copy in B6 (ref. 4). A LOD of 9.3 centred over the *Alad* gene (red curve) represents the differential dosing that occurs between the two strains. The *St7* (NM_022332) gene is differentially spliced at several locations⁵, and, for a stable splice form at the 3' location of the gene, the probe for this gene fortuitously overlapped the region alternatively spliced out in DBA, but not in B6. The differential splicing event is detected by the major QTL (LOD score of 20.1) for *St7*, which is centred over the *St7* gene (blue curve) on chromosome 6. Finally, the *Nnmt* (NM_010924) gene, important for drug metabolism, has been identified as being polymorphic with respect to transcription between the DBA and B6 strains⁶. This polymorphism is confirmed by a major QTL (LOD score of 15.3) for *Nnmt*, centred over the *Nnmt* gene (green curve) on chromosome 9. We note that single nucleotide polymorphisms (SNPs) covered by 60-residue oligonucleotide probes would not be expected to significantly affect transcript abundance measurements among these samples⁷.

Next we investigated a way of combining gene expression, genetics and clinical trait data to elucidate the complexity of common diseases. Major loci controlling complex phenotypes

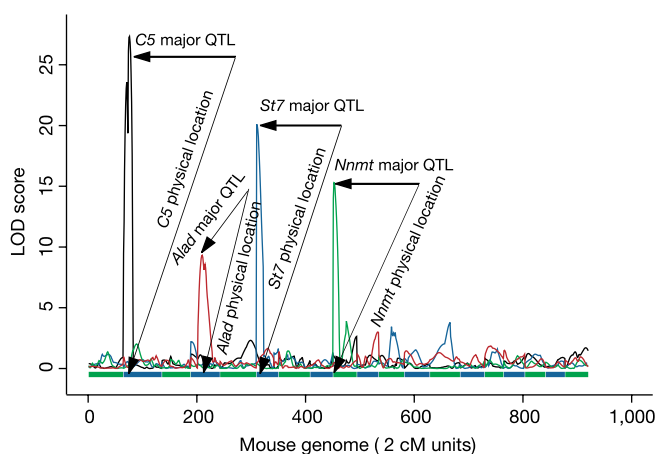


Figure 2 *Cis*-acting eQTL identified for several DNA polymorphisms known to induce transcription polymorphisms. The LOD score curves over the murine autosomal chromosomes are shown for the expression traits of four genes: (1) black curve, *C5* (NM_010406); (2) red curve, *Alad* (AK002300); (3) blue curve, *St7* (NM_022332); and (4) green curve, *Nnmt* (NM_010924). The alternating green and blue strip at the base of the curves represents chromosome boundaries, starting with chromosome 1 and ending with chromosome 19.

such as obesity may potentially affect scores of genes, if not hundreds. We would expect those genes involved in the more downstream aspects of pathways associated with common diseases to have eQTL linked to the major causative loci for those diseases. In addition, there may be heterogeneity among the causative loci for a given disease in a population of interest. When present, this heterogeneity impacts the ability to detect linkages to the causative loci, as the significance of any one locus is diminished when the population is considered as a whole.

Because the mice described above had been on a high-fat, atherogenic diet for 4 months before livers were profiled⁸, they model the spectrum of obesity, diabetes and atherosclerosis in a natural population. The 280 genes depicted in the two-dimensional cluster in Fig. 1b were selected as the most differentially expressed set of genes in mice comprising the upper and lower 25th percentiles of the subcutaneous fat-pad-mass (FPM) trait. When clustering on this set of genes over the high/low FPM mice, the mice cluster almost perfectly into high FPM and low FPM groups (Fig. 1b). Furthermore, there seem to be two distinct expression patterns for mice in the high FPM group, indicating some degree of heterogeneity in the high FPM mice.

Arrows in Fig. 1b highlight five regions in Fig. 1a where more than 50% of the genes in the FPM set genetically link. Taking into account the eQTL distribution for all genes highlighted in Fig. 1a, each of the five hotspot regions defined by the FPM cluster are very significantly enriched for eQTL for the genes in the FPM set. For instance, 25% of the genes in the FPM set have eQTL in a 10-cM region on chromosome 19. This represents an almost fivefold increase over what would be expected by chance, given that roughly 5.5% of genes from the genome-wide gene set have eQTL to this same 10-cM region.

The patterns in Fig. 1b serve to define the obesity trait, FPM, beyond what would be possible without the expression data. There are clearly two distinct patterns associated with high FPM mice. For the FPM trait, a genome-wide scan revealed four classical QTL (clinical QTL or cQTL) with LOD scores greater than 2.0 (ref. 8). To further elucidate this clinical trait, the 111 F₂ animals for which clinical and gene expression data existed were classified into one of the three groups shown in Fig. 1b. Subsequently, separate genetic analyses were performed on two sets of animals: (1) those classified as high FPM group 1 or low FPM; and (2) those classified as high FPM group 2 or low FPM. Figure 1c, d depicts the results of these analyses for two chromosomes. The chromosome 2 FPM QTL was the largest of the four QTL originally identified for this trait, when all animals were considered together. This QTL vanishes when considering the high FPM group 1 with the low FPM group and increases by almost 2 LOD units over the original when considering the high FPM group 2 with the low FPM group. Figure 1d depicts another interesting locus for which the original analysis on the full set of mice yielded no significant QTL for the FPM trait, but where the high FPM group 2 considered with the low FPM group gave rise to a QTL with a significant LOD score, whereas the high FPM group 1 considered with the low FPM group was less significant than that of the full set. These results indicate that the chromosome 2 and 19 QTL each significantly affect only a subset of the F₂ population, a form of heterogeneity that clearly demonstrates the complexity underlying traits such as obesity.

We next focused our efforts on objectively identifying candidate genes for common diseases. An expanded view of the clinical traits and a portion of the gene expression traits linking to the chromosome 2 locus discussed above is given in Fig. 3. Notably, a group of major urinary protein genes represented in the Fig. 1b FPM cluster are linked to the chromosome 2 locus, in addition to seven other loci (all with a LOD score exceeding 2.0), four of which localize together with adiposity or FPM traits. The *Mup1* gene stands out because it was the most highly correlated or had eQTL co-localized with eQTL from many other genes known to be involved in obesity-related

pathways, including retinoid X receptor (Rxr)- γ , Rxr-interacting protein, acyl-coenzyme A oxidase 1, leptin receptor, peroxisome proliferator activated receptor (Ppar)- γ and Lpr6. This demonstrates that the chromosome 2 locus draws together adiposity, FPM, cholesterol and triglyceride levels and is linked to genes with proven roles in obesity and diabetes. Furthermore, the *Mup* genes are members of the lipocalin protein family, and although they are known to have a central role in pheromone-binding processes that affect mouse physiology and behaviour⁹, variations in *Mup* expression have been associated with variations in body weight¹⁰, bone length¹⁰ and levels of very-low-density lipoprotein¹¹.

Most of the genes linked to the chromosome 2 locus do not physically reside on chromosome 2, and so, are at least partially regulated by one or more loci in the chromosome 2 hotspot region. Of the 423 genes linked to the chromosome 2 locus for which we have chromosome mapping information, there are only four eQTL with LOD scores greater than 3.0 that correspond to genes whose physical locations are within 2 cM of the peak. However, only two of the four genes (NM_025575 and NM_15731) had significant genetic interactions with the subcutaneous FPM trait. Gene NM_025575 codes for a dolichyl-diphospho-oligosaccharide-protein glycosyltransferase and gene NM_015731 codes for a cation-transporting ATPase; these genes may be considered as primary causative candidates for the linkage activity at the chromosome 2 locus.

The region supporting the chromosome 2 locus is homologous to human chromosome 20q12-q13.12, a region that has previously been linked to human obesity-related phenotypes^{12,13}. The human orthologues for genes NM_025575 and NM_015731 highlighted in the Fig. 3 reside in the human chromosome 20 region. Although other genes, such as melanocortin 3 receptor (*Mc3r*), have been suggested as possible candidates for obesity at this locus¹³, our data suggest that the genes NM_025575 and NM_015731 may be responsible for the underlying QTL, as discussed above. We note that expression levels of *Mc3r* are not linked to the chromosome 2 locus, and there were no SNPs annotated in the exons or introns of this gene between the C57/BL6 and DBA/2J strains in the most recent build of the Celera RefSNP database, further suggesting that *Mc3r* may not be the chromosome 2 QTL.

To pursue further our survey of the heritability of gene expression, we focused on *Zea mays*, a classical genetic organism that has been studied intensely and where previous reports have

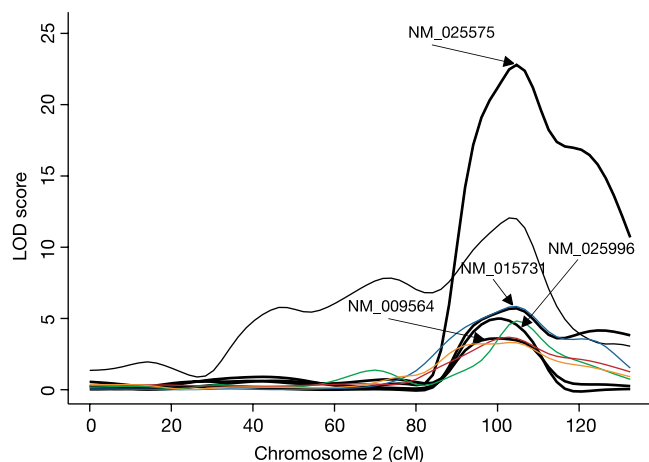


Figure 3 Clinical QTL (cQTL) for obesity-related traits localize together with eQTL. LOD score curves for four obesity-related traits are shown: (1) blue curve, subcutaneous FPM; (2) green curve, perimetrial FPM; (3) red curve, omental FPM; (4) orange curve, adiposity; (5) thin black curve, joint LOD score curve for these four clinical traits considered simultaneously. LOD score curves for four candidate genes that may explain the obesity cQTL are also shown (thick black curves).

demonstrated that protein¹⁴, enzymatic¹⁵ and metabolite¹⁶ levels could be associated with regions of the genome that control complex traits. QTL results from treating gene expression in ear leaf tissue as a quantitative trait in 76 F_2 -derived F_3 progeny constructed from two typical inbred lines of maize—a stiff stalk synthetic type and a Lancaster type—largely parallel those in the mouse described above. Of the 18,805 genes detected as being significantly differentially regulated (type I error = 0.05) in at least 10 samples, there were 6,481 genes with at least one eQTL exceeding a LOD score of 3.0 (with a maximum LOD score of 41.3), and a total of 7,322 QTL overall. Most genes had a single eQTL, and just over 80% of the eQTL exceeding a LOD score of 7.0 were localized together with the physical gene giving rise to the eQTL, when the physical location of the gene could be determined.

Figure 4 represents a new form of genetic interaction. Two genes, each with a significant eQTL on different chromosomes, are seen to have uncorrelated expression values over the 76 samples. However, the genes appear to be interacting by genotype, representing an interaction between genes that is similar to epistasis, but more general because it occurs between the eQTL of different genes. Although interactions such as this (statistically significant at a type I error of 0.05) occurred among fewer than 10% of the genes with significant eQTL (LOD scores greater than 3.0), this form of genetic interaction clearly offers a potentially powerful means to implicate genes as being involved in the same or related pathways. Furthermore, this type of interaction would have been missed by all standard clustering methods applied to the expression data alone, but the relationship was readily detected using statistical genetics models designed to identify interactions among QTL.

As a preliminary survey into the genetics of gene expression in humans, 56 individuals from four CEPH (Centre d'Etude du Polymorphisme Humain) reference families¹⁷ were selected for expression profiling of lymphoblastoid cell lines using a standard human gene oligonucleotide microarray¹⁸. The four families, CEPH/Utah pedigrees 1362, 1375, 1377 and 1408, consisted of large 'sibships' along with parents and grandparents. Heritability analysis was performed for gene expression on a subset of 2,726

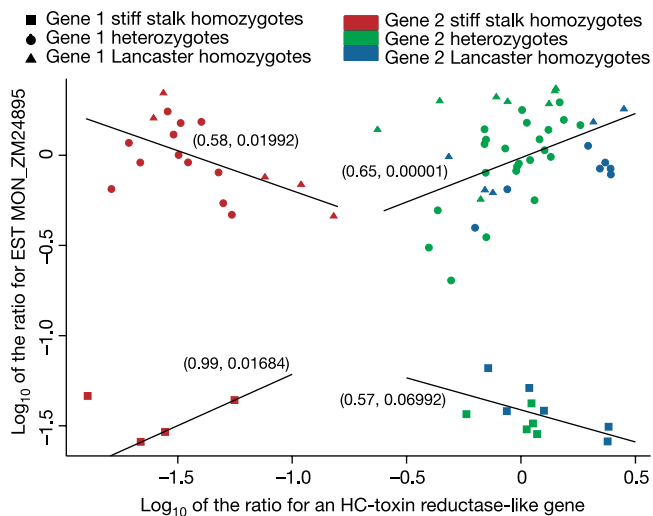


Figure 4 Genes with no overall correlation with respect to expression demonstrate interesting patterns of genetic interaction. The scatter plot shows the mean $\log_{10}$ ratio for two *Zea mays* genes that are uncorrelated overall, each with a significant eQTL (LOD of 24.3 for the gene on the x axis and 24.9 for the gene on the y axis) falling on two separate chromosomes. Patterns are apparent in the plot despite the overall random correlation, as the four groups in each quadrant of the plot are correlated. The least squares regression line is shown for each quadrant, with the correlation coefficient values and corresponding *P*-values given in parentheses. EST, expressed sequence tag. HC, *Helminthosporium carbonum*.

genes that were significantly differentially regulated (type I error = 0.05) within 8 or more of the 16 pedigree founders. We deemed the sample size too small to perform systematic linkage analysis across all genes. We found that for the differentially expressed genes, 29% had a detectable genetic component (type I error < 0.05). This result offers an important glimpse into the genetics of gene expression in humans, with such a large percentage of genes detected with significant heritabilities in such a small sample of 'normal' individuals. We propose that this group of genes may make good therapeutic targets for complex human diseases, given that their degree of genetic control was so readily identifiable in such a small number of families. A complete list of those genes with a significant heritable component and additional information on tests for gender and age effects are provided in the Supplementary Information.

The identification of eQTL for genes expressed in a handful of tissues begins to provide an insight into the genetic networks that underlie the complexity of living systems. The several hundred genes with LOD scores exceeding 20 in the mouse represent a new class of quantitative traits, with linkage significances not commonly seen before in mammalian systems. The causal nature of genetics allows the anchoring of multiple genes under the common control of single or multiple loci, as shown in Fig. 3, thereby providing roots for the graphs that can more completely depict the complicated network of gene interactions at work in complex phenotypes.

The class of genes discussed for Figs 1b and 3 provide objective evidence that many of the genes co-localized to a single QTL hotspot are associated with the obesity-related traits. The patterns of expression serve to refine the obesity phenotype and allow the enrichment of subpopulations that are homogeneous with respect to the underlying causes of obesity in the population. Identifying such subpopulations has significant consequences for drug discovery, as each subpopulation may be more effectively treated by a compound that targets a pathway specifically associated with the disease in that subpopulation.

Combining gene expression, genotype and clinical data in a segregating population has the potential to affect the more significant rate-limiting steps in the drug discovery process: objectively classifying individuals according to disease subtypes and identifying the drivers of the pathways or the causal factors underlying those disease subtypes. In the past, dissecting complex traits using genetics has met with limited success, and until now gene expression has served as an indirect marker for complex traits. We have demonstrated that the combination of gene expression and genetics data has the potential to overcome these barriers. The addition of gene expression data can be used to refine the disease phenotype, directly implicate pathways and genes comprising those pathways associated with the disease phenotype, and identify the key drivers of the pathways underlying the disease phenotype. □

Methods

Preparation of labelled complementary DNA

Lymphoblastoid cell lines from CEPH/Utah pedigree families 1362, 1375, 1377 and 1408 were obtained from Coriell Cell Repositories. Other lymphoblastoid cell lines were established from normal donors by immortalization with Epstein-Barr virus (EBV) as described previously¹⁹. Cells were cultured in RPMI 1640 medium containing 15% fetal bovine serum, and penicillin/streptomycin antibiotics (Invitrogen). Cells were maintained in the log phase of cell growth for at least two days and were collected at densities of $0.4\text{--}0.9 \times 10^6$ cells ml⁻¹.

An F₂ intercross was constructed from C57BL/6J and DBA/2J strains of mice. All mice were housed under conditions meeting the guidelines of the Association for Accreditation of Laboratory Animal Care. Mice were on a rodent chow diet up to 12 months of age, and then switched to an atherogenic high-fat, high-cholesterol diet for another 4 months. More details on this cross are described in ref. 8. Parental and F₂ mice were killed at 16 months of age. At death, the livers were immediately removed, flash-frozen in liquid nitrogen and stored at -80 °C.

An F₃ cross was constructed from standard inbred lines of *Z. mays*: a stiff stalk strain and Lancaster type²⁰. After the initial cross, the F₁ plants were self-crossed to obtain the F₂ progeny, and then the F₂ were self-crossed to obtain the F₃ progeny. The plants were field grown, and at 5 days after flowering, ear leaf tissues were collected from ten F₃ progeny for

each F₂ line. The ear leaf tissues were frozen on dry ice in the field on collection.

The hybridizations for each species were performed in duplicate with fluor reversal. In all cases, total cellular RNA was then purified using an RNeasy Mini kit according to the manufacturer's instructions (Qiagen). For each species, competitive hybridizations were performed by mixing fluorescently labelled antisense RNA (cRNA; 5 µg) from each sample with the same amount of cRNA from a reference pool comprising equal amounts of cRNA from related samples (for more details, see Supplementary Information).

Probe selection for the gene expression arrays

The human microarray contained 24,479 non-control oligonucleotide probes for human genes as described previously¹⁸. The mouse microarray contained 23,574 non-control oligonucleotide probes for mouse genes and 2,186 control probes. Full-length mouse sequences were extracted from murine Unigene clusters, combined with RefSeq mouse sequences and RIKEN full-length sequences. The maize ear leaf microarray carried 24,473 non-control oligonucleotide probes for genes of interest and 1,287 control probes. More than 50,000 non-control probes were selected from maize and rice. These probes were used on ink-jet microarrays to collect expression data for RNA samples from a range of maize tissues. The top-ranking 24,473 probes were carried forward for use on the ear leaf array.

For all microarrays, to select a probe 60 nucleotides in length for each gene sequence, we used a series of filtering steps, taking into account repeat sequences, binding energies, base composition, distance from the 3' end, sequence complexity and potential cross-hybridization interactions⁷. All microarrays used in this study were custom ink-jet microarrays fabricated by Agilent Technologies.

Analysis of expression data

Array images were scanned using the Agilent Dual Laser Microarray scanner (Agilent Technologies) and processed as previously described²¹ to obtain background noise, single-channel intensity and associated measurement error estimates. The colour displays given in panel b of the Fig. 1b show log₁₀ (expression ratio) as: (1) red when the red channel is upregulated relative to the green channel; (2) green when the red channel is downregulated relative to the green channel; (3) black when the log₁₀ (expression ratio) is close to zero; and (4) grey when data from one or both of the channels for a given probe is unreliable.

Linkage and data analysis

Variance components analysis²² was used to estimate the heritability of gene expression, as measured by the mean log₁₀ expression ratio, for each of the 2,726 mRNAs that were significantly differentially expressed in the founders, and to test whether the heritability was significantly different from zero. Genes were defined as differentially regulated if eight or more founders had a *P*-value for differential expression less than 0.05.

A complete linkage map for all chromosomes except Y in mouse was constructed at an average density of 13 cM using microsatellite markers. Linkage maps were constructed and QTL analysis was performed using MapMaker QTL²³ and QTL Cartographer²⁴. Log of the odds ratio (LOD) scores were calculated at 2-cM intervals throughout the genome for each of the 23,574 genes represented on the mouse microarray.

A complete linkage map for all chromosomes in *Z. mays* was constructed at an average density of 12 cM using microsatellite markers. Linkage maps were constructed based on genotypes from the F₂ progeny using MapMaker QTL. Mapping of QTL for gene expression traits was carried out as described previously²⁵. For the interaction result presented in Fig. 4, individual specific LOD score vectors were calculated for each QTL, and correlations were computed between all vector pairs. See the Supplementary Information for more details on the segregation and linkage analyses performed over all species.

Received 7 October 2002; accepted 10 January 2003; doi:10.1038/nature01434.

- Eaves, I. A. *et al.* Combining mouse congenic strains and microarray gene expression analyses to study a complex trait: the NOD model of type 1 diabetes. *Genome Res.* **12**, 232–243 (2002).
- Karp, C. L. *et al.* Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nature Immunol.* **1**, 221–226 (2000).
- Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
- Claudio, L., Lee, T., Wolff, M. S. & Wetmur, J. G. A murine model of genetic susceptibility to lead bioaccumulation. *Fundam. Appl. Toxicol.* **35**, 84–90 (1997).
- Huang, Y. H., Chen, Y. T., Lai, J. J., Yang, S. T. & Yang, U. C. PALS db: Putative Alternative Splicing database. *Nucleic Acids Res.* **30**, 186–190 (2002).
- Yan, L., Otterness, D. M., Kozak, C. A. & Weinshilboum, R. M. Mouse nicotinamide N-methyltransferase gene: molecular cloning, structural characterization, and chromosomal localization. *DNA Cell Biol.* **17**, 659–667 (1998).
- Hughes, T. R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnol.* **19**, 342–347 (2001).
- Drake, T. A. *et al.* Genetic loci determining bone density in mice with diet-induced atherosclerosis. *Physiol. Genomics* **5**, 205–215 (2001).
- Timm, D. E., Baker, L. J., Mueller, H., Zidek, L. & Novotny, M. V. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Sci.* **10**, 997–1004 (2001).
- Metcalfe, D. *et al.* Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature* **405**, 1069–1073 (2000).
- Swift, L. L., Valyi-Nagy, K., Rowland, C. & Harris, C. Assembly of very low density lipoproteins in mouse liver: evidence of heterogeneity of particle density in the Golgi apparatus. *J. Lipid Res.* **42**, 218–224 (2001).
- Borecki, I. G., Rice, T., Perusse, L., Bouchard, C. & Rao, D. C. An exploratory investigation of genetic linkage with body composition and fitness phenotypes: the Quebec Family Study. *Obesity Res.* **2**, 213–219 (1994).

13. Lemberas, A. V. *et al.* Identification of an obesity quantitative trait locus on mouse chromosome 2 and evidence of linkage to body fat and insulin on the human homologous region 20q. *J. Clin. Invest.* **100**, 1240–1247 (1997).
14. Damerval, C., Maurice, A., Josse, J. M. & de Vienne, D. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**, 289–301 (1994).
15. Causse, M. *et al.* Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. *Mol. Breed.* **1**, 259–272 (1995).
16. Byrne, P. F. *et al.* Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl Acad. Sci. USA* **93**, 8820–8825 (1996).
17. Dausset, J. *et al.* Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577 (1990).
18. van't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
19. Tosatio, G. Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines. *Curr. Protoc. Immunol.* **1**, 7.22.1–7.22.3 (1991).
20. Troyer, A. F. Background of US hybrid corn. *Crop Sci.* **39**, 601–626 (1999).
21. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
22. Amos, C. I. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**, 535–543 (1994).
23. Lincoln, S. E., Daly, M. J. & Lander, E. S. *MAPMAKER/QLT User's Manual* (Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 1993).
24. Basten, C. A., Weir, B. S. & Zeng, Z. B. *QLT Cartographer User's Manual* (Department of Statistics, North Carolina State University, Raleigh, 1999).
25. Fisch, R. D., Ragot, M. & Gay, G. A generalization of the mixture model in the mapping of quantitative trait loci for progeny from a biparental cross of inbred lines. *Genetics* **143**, 571–577 (1996).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements Rosetta Inpharmatics is a wholly owned subsidiary of Merck & Co., Inc.

Competing interests statement The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to E.E.S.
(e-mail: eric_schadt@merck.com) or S.H.F. (e-mail: stephen_friend@merck.com).

Glycine binding primes NMDA receptor internalization

Yi Nong^{*†‡}, Yue-Qiao Huang^{*†‡}, William Ju^{*§}, Lorraine V. Kalia^{*||},
Gholamreza Ahmadian^{*§}, Yu Tian Wang^{*§¶} & Michael W. Salter^{*†||}

^{*} Programme in Brain and Behaviour, Hospital for Sick Children, and
[†] Department of Physiology, [§] Department of Pathobiology and Laboratory
Medicine, and ^{||} Institute of Medical Science, University of Toronto, Toronto,
Ontario M5G 1X8, Canada

[¶] Department of Medicine and Brain Research Centre, Vancouver Hospital and
Health Sciences Centre, University of British Columbia, Vancouver, British
Columbia V6T 1Z3, Canada

[‡] These authors contributed equally to this work

NMDA (N-methyl-D-aspartate) receptors (NMDARs) are a principal subtype of excitatory ligand-gated ion channel with prominent roles in physiological and disease processes in the central nervous system¹. Recognition that glycine potentiates NMDAR-mediated currents² as well as being a requisite co-agonist of the NMDAR subtype of 'glutamate' receptor³ profoundly changed our understanding of chemical synaptic communication in the central nervous system. The binding of both glycine and glutamate is necessary to cause opening of the NMDAR conductance pore¹. Although binding of either agonist alone is insufficient to cause current flow through the channel, we report here that stimulation of the glycine site initiates signalling through the NMDAR complex, priming the receptors for clathrin-dependent endocytosis. Glycine binding alone does not cause the receptor to

be endocytosed; this requires both glycine and glutamate site activation of NMDARs. The priming effect of glycine is mimicked by the NMDAR glycine site agonist D-serine, and is blocked by competitive glycine site antagonists. Synaptic as well as extrasynaptic NMDARs are primed for internalization by glycine site stimulation. Our results demonstrate transmembrane signal transduction through activating the glycine site of NMDARs, and elucidate a model for modulating cell–cell communication in the central nervous system.

The turnover of cell surface NMDARs is much slower than that of other ligand-gated ion channels, such as AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)^{4–9} and GABA_A (γ -aminobutyric acid A)¹⁰ receptors, and NMDARs are often considered 'stable' in the plasma membrane. However, recent evidence has raised the possibility that NMDARs may undergo regulated transport to and from the cell surface^{11–13}. Here, we recorded whole-cell currents from neurons acutely isolated from the CA1 region of the hippocampus. We probed NMDAR activity with regularly timed test applications of NMDA (50 μ M) plus glycine (1 μ M), and observed that the NMDAR-mediated currents were stable for recording periods of 45 min or longer (Fig. 1a). In contrast, on administering a conditioning stimulus for NMDARs (1 mM NMDA plus 100 μ M glycine; 2–3 min duration) there was a sharp decline in the NMDAR currents followed by a progressive decrease to a level $57 \pm 3\%$ (mean $\pm$ s.e.m.; $n = 6$ cells) of the control NMDAR current amplitude (Fig. 1b). The depression of NMDAR currents was stable and the currents did not recover even during the longest recordings, up to 1 h. The depression of NMDAR currents was associated with no change in reversal potential, no loss of neuronal integrity and no alteration in single channel conductance or mean open time (see Supplementary Information 1).

It is possible that the decline in NMDAR currents was due to a decrease in the number of cell-surface receptors. Thus, we determined whether the depression of NMDAR currents might be due to clathrin-dependent receptor internalization. For this we used an amphiphysin SH3 domain recombinant protein^{6,14}, which prevents recruitment of dynamin to clathrin-coated pits and thereby blocks endocytosis. We found that when wild-type amphiphysin SH3 domain protein was administered intracellularly during whole-cell recording, the co-agonist conditioning stimulation produced a small, transient decline in NMDAR current amplitude but no sustained decrease in the currents (Fig. 1c). In contrast, when we administered an amphiphysin SH3 domain protein with a double point mutation (G684R, P687L)¹⁵ that renders it incapable of blocking endocytosis, the co-agonist stimulation caused a sustained decline in NMDAR currents (Fig. 1c) that was not different from that in control recordings.

To determine directly whether conditioning of NMDARs leads to loss of the receptors from the cell surface as a consequence of clathrin-dependent receptor endocytosis, we used a cell enzyme-linked immunosorbent assay (ELISA) (Fig. 1d) with hippocampal neurons in primary culture. Using control extracellular bathing solution we found that there was little internalization of NMDARs during the 15-min experimental period (Fig. 1d). On the other hand, exposing the cells to the conditioning stimulus for 5 min caused $27 \pm 3\%$ ($n = 6$) of the pre-labelled NMDARs to be internalized over the subsequent 10-min period (Fig. 1d). To test whether the internalization of NMDARs required clathrin assembly we used extracellular hypertonic sucrose solution, known to block clathrin-mediated endocytosis¹⁶, which was found to prevent the internalization (Fig. 1d, top). NMDAR internalization was produced when the concentration of NMDA in the co-agonist conditioning stimulus was reduced to 50 μ M (Fig. 1d, bottom). In contrast, neither the application of NMDA nor glycine alone caused internalization of NMDARs. Furthermore, applying the competitive NMDAR antagonist AP5 (D-2-amino-5-phosphonopivalic acid)¹⁷—which binds to the glutamate-binding site but does not

activate the receptors—together with glycine did not cause NMDAR internalization. These results indicate that co-agonist activation of NMDARs provokes their robust internalization by clathrin-dependent endocytosis.

When we applied AP5 together with the conditioning NMDA and glycine stimulus during electrophysiological recordings, we found that the test response immediately after the conditioning stimulus had the same amplitude as did the control responses before conditioning (Fig. 2a). Thus, AP5 eliminated the immediate sharp decrease in NMDAR currents produced by the conditioning stimulus, in accord with NMDAR antagonism by AP5 and the lack of effect of AP5 plus glycine on NMDAR internalization. Notably, however, the amplitude of subsequent test responses was not

maintained and the responses progressively declined, reaching a level $58 \pm 8\%$ of the pre-conditioning control ($n = 4$ cells; Fig. 2a). This decrease in NMDAR amplitude was not different from that produced with the conditioning stimulus lacking AP5. Conditioning with AP5 alone did not cause any change in subsequent test responses. Thus, the conditioning stimulus provoked a sharp decline that was dependent on NMDAR activation during this stimulus, as indicated by the AP5 blockade, and produced a progressive decline that was not prevented by AP5.

Because AP5 is competitive with glutamate at its binding site on the NMDAR, we hypothesized that induction of the progressive decline in NMDAR responses may not require binding to this site but rather may depend on activating the glycine-binding site. This was tested by conditioning with glycine, without NMDA, which caused NMDAR test responses to decline progressively to a stable level at $63.5 \pm 2\%$ ($n = 8$ cells) of the control level (Fig. 2b), which was not different from that reached after the NMDA (1 mM) plus glycine (100 μ M) conditioning stimulus, with or without AP5. We found that D-serine (30 μ M), another endogenous agonist of the glycine site of the NMDAR¹⁸, also produced a progressive decline in NMDAR test responses to a level not different from that reached after conditioning with glycine (Fig. 2b). Moreover, the effect of

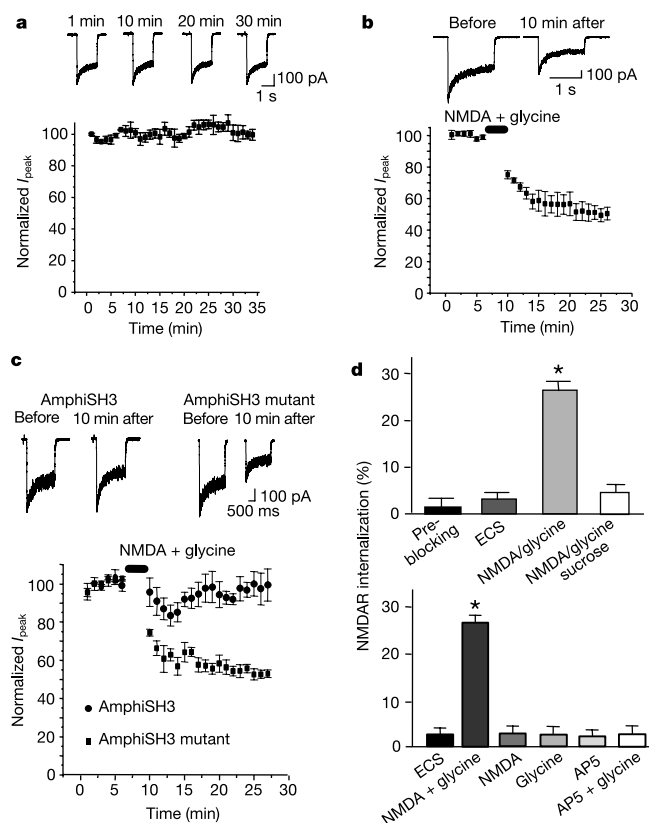


Figure 1 Co-agonist stimulation induces dynamin-dependent NMDA receptor internalization. **a**, Plot of the normalized NMDAR peak currents (mean $\pm$ s.e.m.) recorded in acutely isolated rat hippocampal CA1 neurons ($n = 8$ cells). Peak currents were normalized to the average of the first six responses evoked by test applications of NMDA (50 μ M) and glycine (1 μ M) in this and subsequent figures. Traces (top) show responses to the test applications recorded from one cell at the times indicated. **b**, Normalized NMDAR peak currents ($n = 6$ cells) are plotted before and after conditioning with NMDA (1 mM) plus glycine (100 μ M) applied during the time indicated. Shown above are traces of NMDA-evoked whole-cell currents before and 10 min after conditioning stimulation. **c**, Normalized peak NMDA-evoked currents from recordings with intracellular administration of amphiphysin SH3 domain (amphiSH3): wild type (circles; $n = 8$ cells) or G684R, P687L mutant (squares; $n = 5$ cells) protein. **d**, The top panel shows NMDAR internalization (mean $\pm$ s.e.m.; percentage of total) as quantified by cell ELISA assay using hippocampal neurons in primary culture. Cultures ($n = 12$ in three separate experiments) were treated with control ECS or ECS containing NMDA (1 mM) plus glycine (100 μ M) with or without hypertonic sucrose (0.45 M). For pre-blocking conditions cultures were not treated with ECS. The bottom panel shows quantification of NMDAR internalization for cultures treated with control ECS or ECS supplemented with NMDA (50 μ M) plus glycine (100 μ M), NMDA (50 μ M), glycine (100 μ M), AP5 (50 μ M), or glycine (100 μ M) plus AP5 (50 μ M) ($n = 18$ in three separate experiments).

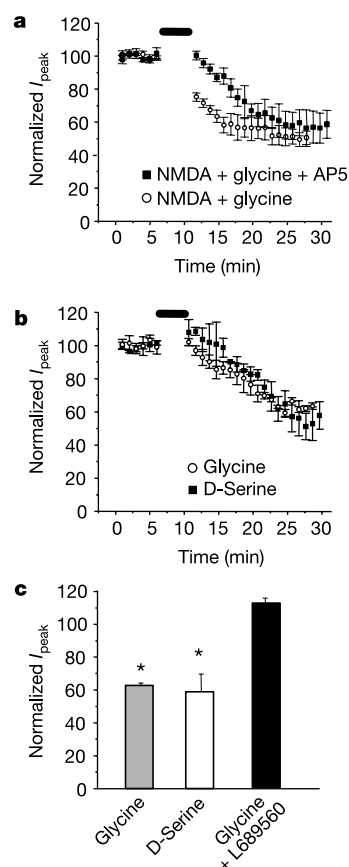


Figure 2 Glycine site stimulation initiates a progressive decline in NMDA receptor-mediated currents. **a**, Normalized peak NMDA-evoked test currents are plotted. The conditioning stimulus of NMDA (1 mM) plus glycine (100 μ M) without (open circles; $n = 5$ cells) or with (filled squares; $n = 4$ cells) AP5 (500 μ M) was applied during the period indicated by the black bar. **b**, Normalized peak NMDA currents during recordings in which the conditioning stimulus was ECS supplemented with glycine (100 μ M, open circles; $n = 8$ cells) or D-serine (30 μ M, filled squares; $n = 5$ cells). **c**, Histogram showing averaged, normalized peak NMDA currents 20 min after conditioning with glycine (100 μ M; $n = 8$ cells), D-serine (30 μ M; $n = 5$ cells), or glycine (100 μ M) plus L689560 (1 μ M; $n = 5$ cells).

conditioning with glycine was eliminated by co-administering the glycine site antagonist L689560 (refs 19, 20) (Fig. 2c) but not by AP5 (Supplementary Information 2). From these results we conclude that the progressive decline in NMDAR currents was initiated by stimulating the glycine site, independent of the glutamate site, of the NMDAR during the conditioning stimulation.

To determine whether glycine stimulation initiated a slowly developing process in the cells that was independent of NMDAR activity or whether subsequent NMDAR activation was required to produce the decline in the responses, we performed a separate series of experiments in which the start of the test applications was delayed 5 min after the end of the conditioning glycine stimulus. We found that the first NMDAR response after the 5-min delay was not different in amplitude from the test responses immediately in the control period preceding the conditioning stimulation (Fig. 3a; average $97 \pm 2\%$ control, $n = 5$ cells). However, the NMDAR responses declined with a time course that paralleled that of the

decline of responses when the test applications were started without any delay period (Fig. 3a). Thus, rather than causing the depression of the NMDAR responses, glycine primed the NMDARs for depression, which was subsequently produced by stimulating both co-agonist sites. The priming of depression of NMDAR currents was observed with native receptors in neurons as well as with recombinant NMDARs expressed heterologously (Supplementary Information 3).

These findings raise the possibility that extracellular binding of glycine to NMDARs causes a transmembrane signalling process by means of the receptors that results in intracellular biochemical alterations. We explored this possibility by examining the association between the NMDARs and the AP2 adaptor protein complex: a principal component of the intracellular endocytic machinery²¹. Using proteins prepared under weakly denaturing conditions from acute hippocampal slices—conditions that preserve many protein–protein interactions—we immunoprecipitated the AP2 complex with an antibody against a common subunit of the complex, the adaptin $\beta 2$ subunit, and probed for the level of co-immunoprecipitating NMDARs. Under control conditions we found a low level of co-precipitation of NMDARs together with adaptin $\beta 2$ (Fig. 3b–d); this binding was not detected previously when binding between AMPA receptor subunit GluR2 and adaptin $\beta 2$ was evident⁶, indicating that the level of basal association between AP2 and NMDARs is much less than that between AP2 and AMPA receptors. However, in slices treated with glycine (100 μ M) or glycine (100 μ M) plus AP5 (100 μ M), the level of NMDARs precipitating together with adaptin $\beta 2$ increased greatly, indicating that glycine enhanced the association of AP2 and NMDARs (Fig. 3b–d). In contrast, treating slices with NMDA (50 μ M) or NMDA (50 μ M) plus glycine (1 μ M) produced no change in the adaptin $\beta 2$ –NMDAR association (Fig. 3b). Moreover, treating slices with 1 mM NMDA also produced no change in the adaptin $\beta 2$ –NMDAR association (Fig. 3c). The increase in precipitation of NMDARs together with adaptin $\beta 2$ produced by glycine was prevented by L689560 or another glycine site antagonist MDL29951 (refs 20, 22) (Fig. 3d), which themselves had no effect on the association of AP2 and NMDARs. Thus, stimulation of the extracellular glycine site results in an intracellular process, or processes, that leads to recruitment of the endocytic machinery to NMDARs; however, this recruitment, in and of itself, does not cause NMDAR internalization.

To determine whether priming NMDARs with glycine readies the receptors for subsequent internalization on activation, we compared the effect of receptor activation by NMDA (50 μ M) plus glycine (1 μ M), identical to the co-agonist concentrations used to evoke test NMDAR whole-cell currents, with and without pre-treatment with glycine (100 μ M, 5 min). Without glycine pre-treatment, we found that activating the receptors with NMDA (50 μ M) plus glycine (1 μ M) did not cause them to be internalized (Fig. 4a). However, in neurons that were pre-treated with glycine, subsequent receptor activation did provoke robust NMDAR internalization. The internalization was prevented when receptor activation was blocked by including AP5 with the NMDA and glycine treatment (Fig. 4a), and internalization was produced when NMDA was replaced by the endogenous transmitter glutamate (Fig. 4a). Thus, the internalization process itself was dependent on NMDAR activation. As glycine on its own does not provoke NMDAR internalization (Figs 1d and 4a) our findings indicate that glycine treatment primes, but does not induce, NMDAR endocytosis, and that the receptor internalization process itself is dependent on co-agonist activation of NMDARs.

We next sought to determine whether priming was dependent on stimulating the glycine site of NMDARs. Applying the glycine site antagonist L689560 together with glycine during pre-treatment prevented the co-agonist treatment from producing NMDAR internalization (Fig. 4b). Similarly, L689560 pre-treatment on its

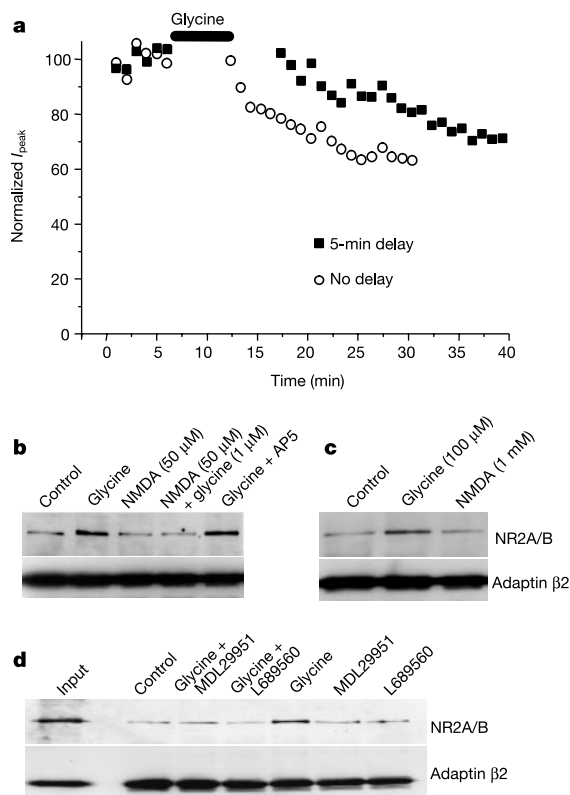


Figure 3 Glycine binding primes the decline of NMDA responses and recruits adaptin $\beta 2$ to the NMDA receptor complex. **a**, Peak NMDA-evoked test currents plotted for two cells conditioned with glycine (100 μ M) during the period indicated. For one cell, test applications were re-started immediately after the glycine conditioning (open circles). For the other cell, there was a delay of 5 min before re-starting the test applications (filled squares). **b**, Co-immunoprecipitation of NMDARs and adaptin $\beta 2$ from acute hippocampal slices treated for 5 min with control ECS, glycine (100 μ M), NMDA (50 μ M), NMDA (50 μ M) plus glycine (1 μ M), or glycine (100 μ M) plus AP5 (100 μ M). Proteins were immunoprecipitated with anti-adaptin $\beta 2$, resolved by SDS–PAGE, and analysed by immunoblotting with anti-NMDAR2A/B (top panel). The blot was then re-probed with anti-adaptin $\beta 2$ (bottom panel). Similar results were observed in each of three separate experiments. **c**, Co-immunoprecipitation of NMDARs and adaptin $\beta 2$ from slices treated for 5 min with control ECS, glycine (100 μ M) or NMDA (1 mM). Shown is the representative of three similar experiments. **d**, Co-immunoprecipitation of NMDARs and adaptin $\beta 2$ from slices treated for 5 min with control ECS, glycine (100 μ M) with or without L689560 (1 μ M) or MDL29951 (1 μ M), or with L689560 or MDL29951 alone. In the lane marked input, 50 μ g of protein without immunoprecipitation was loaded. Similar results were obtained in each of three experiments.

own did not permit subsequent application of co-agonists to stimulate internalization. Thus, glycine site activation is required to prime NMDAR internalization but glycine site occupation with a non-activating ligand is not sufficient to prime the receptors. The effector concentration for half-maximum response (EC_{50}) for the priming effect of glycine was $39 \pm 2 \mu\text{M}$ (Fig. 4c).

To examine whether NMDARs were primed for internalization through the selective activation of the glutamate-binding site, we pre-treated neurons for 5 min with NMDA ($100 \mu\text{M}$ or 1 mM). The pre-treatment was carried out with co-application of L689560 to eliminate any activation of the glycine site. After subsequent treatment with NMDA ($50 \mu\text{M}$) and glycine ($1 \mu\text{M}$), there was no difference in NMDAR internalization compared to pre-treatment with extracellular solution (ECS; Fig. 4b). Thus, activating the glutamate site, in contrast to the glycine site, is not sufficient to prime NMDARs for internalization.

To determine whether the internalization of NMDARs after glycine priming was mediated through clathrin-dependent endocytosis, we made use of a dynamin inhibitory peptide, QVPSRPNRAP, which competitively binds with the amphiphysin SH3 domain¹⁵ and prevents endocytosis when administered intracellularly^{5,10}. Myristoylated QVPSRPNRAP (Myr-4-QVPSRPNRAP) is a membrane-permeant form of the peptide that prevents endo-

cytosis²³. We found that applying Myr-4-QVPSRPNRAP in bath solution inhibited glycine-primed internalization of NMDARs (Fig. 4d). As a negative control we applied the non-myristoylated QVPSRPNRAP, which is not membrane permeant, to bath solution and found that it did not affect the internalization of NMDARs (Fig. 4d). These findings indicate that stimulation of the glycine site alone does not cause the process of NMDAR internalization but rather primes the receptors for clathrin-dependent internalization, which is then produced as a consequence of co-agonist stimulation of the receptor.

In neurons of the central nervous system (CNS), NMDARs are localized at glutamatergic synapses and also at extrasynaptic sites. To investigate directly whether synaptic NMDA receptors are primed by glycine and subsequently internalized, we studied miniature excitatory postsynaptic currents (mEPSCs) during whole-cell recording from hippocampal neurons in primary culture. These mEPSCs show NMDAR- and AMPAR-mediated components²⁴. We found that the NMDAR component of the mEPSCs decreased to $61 \pm 4\%$ of the initial level ($n = 10$ cells, Fig. 4e, g) after bath application of glycine ($100 \mu\text{M}$, 5 min) followed by NMDA ($50 \mu\text{M}$) plus glycine ($1 \mu\text{M}$, 5 min). In contrast, during control recordings without glycine and NMDA plus glycine treatment, the NMDAR component of mEPSCs was stable (Fig. 4g). Intracellular appli-

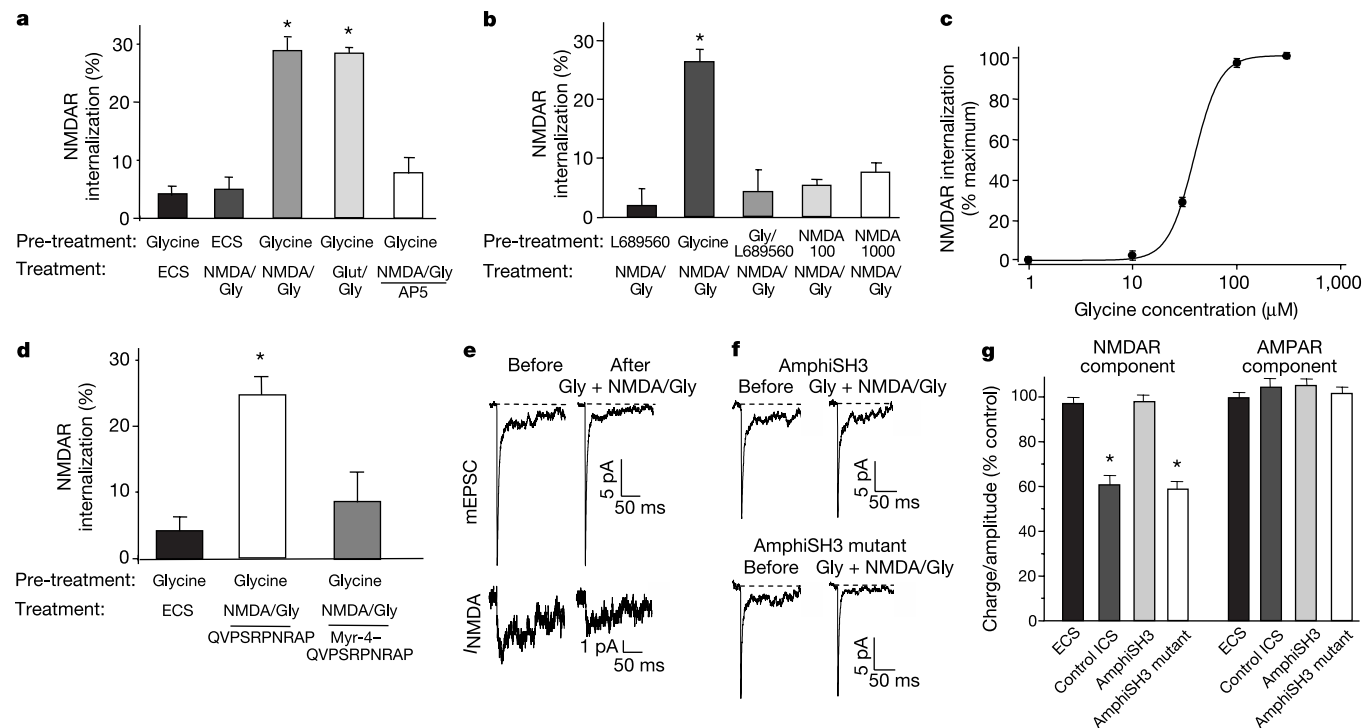


Figure 4 Glycine primes dynamin-dependent internalization of NMDA receptors and leads to preferential suppression of NMDAR- but not AMPAR-mediated synaptic currents.

a, Cultures were pre-treated with ECS or glycine ($100 \mu\text{M}$). Next, cultures were treated with ECS or with NMDA ($50 \mu\text{M}$) plus glycine ($1 \mu\text{M}$) with or without AP5 ($50 \mu\text{M}$), or with glutamate ($50 \mu\text{M}$) plus glycine ($1 \mu\text{M}$) ($n = 18$). NMDAR internalization was quantified by cell ELISA from hippocampal cultures subject to pre-treatment/treatment models in **a–d**. **b**, Pre-treatment (5 min) was as indicated (NMDA $100 \mu\text{M}$ or $1,000 \mu\text{M}$). Treatment: NMDA ($50 \mu\text{M}$) plus glycine ($1 \mu\text{M}$, 5 min; $n = 18$). **c**, Concentration-dependence of the glycine priming effect. Each data point is the mean $\pm$ s.e.m. of NMDAR internalization subtracted from basal NMDAR internalization at each concentration of glycine ($n = 6$ for each point), normalized to the maximum internalization. The solid curve is the best fit of the data to the logistic equation $E = E_{\text{max}}/(1 + (EC_{50}/[\text{glycine}])^n)$; where EC_{50} was $39 \mu\text{M}$ and n was 3.5. **d**, Histogram showing experiments ($n = 6$) in which a membrane-permeant form of the dynamin inhibitory peptide (Myr-4-QVPSRPNRAP; $50 \mu\text{M}$) or the

membrane non-permeant form (QVPSRPNRAP; $50 \mu\text{M}$) were included throughout.

e, Representative averaged mEPSCs (top traces) or NMDAR components (I_{NMDA}) (bottom traces) obtained before and 10 min after pre-treatment with glycine ($100 \mu\text{M}$) and treatment with NMDA ($50 \mu\text{M}$) plus glycine ($1 \mu\text{M}$, 5 min). **f**, Traces are representative average mEPSCs obtained before and 10 min after pre-treatment with glycine and treatment with NMDA plus glycine. Amphiphysin SH3 domain wild-type (top) or mutant (bottom) fusion protein was included in the recording pipette throughout the experiment. **g**, Mean values in NMDAR components and AMPAR components of averaged mEPSCs after treatment with NMDA plus glycine, expressed as percentage of control. Recordings were done with ECS pre-treatment followed by NMDA plus glycine (ECS; $n = 5$ cells) as indicated; in all other experiments, cells were pre-treated with glycine ($100 \mu\text{M}$). The intracellular solution was control (Control ICS; $n = 10$ cells), or included amphiphysin SH3 domain fusion protein (AmphiSH3; $n = 12$ cells) or mutant SH3 domain protein (AmphiSH3 mutant; $n = 8$ cells). Asterisk, $P < 0.005$, t -test, versus ECS.

cation of the wild-type amphiphysin SH3 domain fusion protein during whole-cell recordings prevented the decrease of the NMDAR component of the mEPSCs caused by the glycine followed by NMDA plus glycine treatment ($n = 12$ cells, Fig. 4f, g). However, when the mutant amphiphysin SH3 domain was administered into the cells, the NMDAR component of the mEPSCs declined to $59 \pm 3\%$ of the initial level ($n = 8$ cells, Fig. 4f, g). In contrast to the NMDAR component, the AMPAR component of the mEPSCs was unaffected by treating with glycine followed by NMDA plus glycine (Fig. 4e–g). Together, these results indicate that glycine stimulation differentially primes synaptic NMDARs, but not synaptic AMPA receptors, for internalization.

Our results show that NMDARs undergo regulated internalization on co-agonist stimulation, with a separable priming process initiated by glycine binding. A principal consequence of NMDAR internalization is that the receptors are removed from the cell surface, and are thereby prevented from further activation by extracellular ligands. Thus, neuronal processes that are dependent on NMDAR activation will be suppressed by the regulated internalization of the receptors. Internalized NMDARs might be recycled to the cell surface, be targeted for degradation or be engaged by intracellular biochemical cascades.

The priming that we have discovered is a potential mechanism to explain the basal stability of NMDARs on the cell surface. Under resting conditions in most regions of the CNS, NMDARs may be protected from regulated internalization because the basal extracellular concentration of glycine²⁰ is just below the 'set point' of the internalization priming mechanism. The priming mechanism may become engaged to permit NMDAR internalization when extracellular glycine concentration is greater than the set point: this may occur in the cerebellum or spinal cord where basal concentrations of glycine greater than $20 \mu\text{M}$ have been reported; with high levels of neuronal firing activity either physiologically or pathologically, for example during seizures; or in pathological conditions such as CNS ischaemia or glycine encephalopathy, where extracellular glycine concentrations in the tens to hundreds of micromolars are observed²⁰. Also, because D-serine can produce the priming effect, it is important to consider not only the glycine concentration but also that of glycine and D-serine combined. Various mechanisms, such as allosteric regulation by polyamines, Mg^{2+} and Ca^{2+} , may cause marked changes in potency at the glycine site²⁰. Thus, the priming mechanism might also become engaged if the set point for priming is lowered such that, even without increasing the basal level, the extracellular concentration of glycine and serine might be sufficient to produce the priming effect. We estimate that glycine is about 20–100 times more potent as a co-agonist mediating NMDAR gating as compared with the priming effect discovered here. We suggest two potential explanations for the differing potencies of glycine for these two effects. One is that there might be two different affinities of binding at the glycine site^{25–27} (possibly one coupled to gating and the other coupled to priming), and the second is that the "coupling gain"²⁸ for the coupling between glycine binding and channel gating may be different from that for the coupling between glycine binding and priming (see Supplementary Information 4).

Glycine, without co-agonist activation of the glutamate site, does not evoke NMDAR currents, and thus our findings demonstrate that the priming effect of glycine and the recruitment of AP2 to the receptor complex do not require current flow through NMDARs. A core concept about NMDARs is that their primary signalling mechanisms, depolarization and influx of Ca^{2+} are consequences of transmembrane current flow through the receptor, although it has been suggested that NMDARs may signal independently of ionic flux¹². From the results of our study, we propose a new concept for signalling by NMDARs: transmembrane signal transduction by selectively stimulating the glycine, but not the glutamate, binding site on the receptors. Hence, glycine functions as more than simply a

co-agonist for NMDARs, and the glycine and glutamate binding sites of the receptor can have divergent functions. As NMDARs are implicated in a diversity of physiological and pathological processes throughout the central nervous system¹, glycine priming of NMDAR internalization may have a general role in regulating nervous system functioning in normal and disease states. □

Methods

CA1 neuron isolation and whole-cell patch clamp recording

CA1 neurons were isolated from hippocampus slices taken from postnatal Wistar rats (17–25 days) using described procedures²⁹. Only the neurons that retained their pyramidal shape were used for recordings. Whole-cell recordings were made at room temperature ($20\text{--}22^\circ\text{C}$). After formation of a whole-cell configuration, the recorded neurons were voltage-clamped (-60 mV) and lifted into the stream of solution supplied by a computer-controlled, multi-barrelled fast perfusion system. The extracellular solution was composed of 140 mM NaCl , 1.3 mM CaCl_2 , 5 mM KCl , 25 mM HEPES , 33 mM glucose and $500\text{ nM tetrodotoxin}$, with pH 7.35 and osmolarity 330 mosM . The intracellular solution contained 140 mM CsF , 10 mM BAPTA , 1 mM CaCl_2 , 2 mM MgCl_2 , 10 mM HEPES and $4\text{ mM K}_2\text{ATP}$, with pH 7.3 and osmolarity 300 mosM . NMDAR currents were recorded using an Axopatch 1-D amplifier, data were digitized with DigiData 1200A, filtered (2 kHz), and acquired by the pClamp8.1 program. Recordings in which the series resistance varied by more than 10% were rejected.

Cell ELISA assay

Assays were carried out essentially as described⁶. Briefly, hippocampal neurons were cultured in 12-well plates (approximately 1.5×10^5 cells per well). After removing the medium, neurons were placed into extracellular solution (ECS) and cooled to 4°C to inhibit receptor transport. Cells were then incubated with anti-NMDAR1 monoclonal antibody against the extracellular amino terminus of mouse NMDAR1 (Chemicon; $2.5\text{ }\mu\text{g ml}^{-1}$) for 30 min at 4°C for the purpose of pre-labelling NMDA receptors on the cell surface. For experiments in Fig. 1, cells were treated at 37°C for 5 min with ECS or with NMDA and/or glycine and/or AP5 in ECS at concentrations indicated in the legend; after this treatment period the cells were washed with control ECS for 10 min before fixation. ECS was replaced where indicated with ECS containing hypertonic sucrose (0.45 M). For experiments in Fig. 4, cells were pre-treated at 37°C for 5 min with ECS or with glycine, NMDA and/or L689560 in ECS, at concentrations indicated in the legend. Cells were then treated for 5 min with control ECS or with $50\text{ }\mu\text{M}$ NMDA and $1\text{ }\mu\text{M}$ glycine followed by incubating for 5 min with control ECS before fixation. AP5 or the dynamin peptides were added to the treatment solution as indicated. NMDA in the treatment solution was replaced, as indicated, by glutamate ($50\text{ }\mu\text{M}$). Cells were fixed for 10 min with 4% paraformaldehyde in PBS, and then half of the samples were permeabilized with 0.1% Triton X-100 in PBS for 5 min. The cells were then rinsed with PBS and incubated for 1 h at room temperature with secondary mouse antibody conjugated to horseradish peroxidase (HRP) ($1:1,000$, Amersham). After washing with D-PBS, HRP substrate OPD was added to produce a colour reaction that was stopped with 0.2 volume of 3N HCl . The optical density of 1 ml of supernatant was read on a spectrophotometer at 492 nm . The rates of cell surface expression of NMDA receptors were presented as the ratio of colorimetric readings under non-permeabilized conditions versus those under permeabilized conditions. Analysis was done using at least six separate dishes in each group.

Immunoprecipitation of AP2 with NMDARs

Hippocampal slices were prepared from postnatal rats (Wistar 20–24 days) as described²⁹. Six slices ($400\text{ }\mu\text{m}$) were treated with ECS containing various drugs such as glycine agonists or antagonists for 5 min and then pooled. All the ECS contain strychnine ($1\text{ }\mu\text{M}$), bicuculline ($20\text{ }\mu\text{M}$) and tetrodotoxin ($1\text{ }\mu\text{M}$). The tissue was then homogenized in ice-cold lysis buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl , 2 mM EDTA , 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, protease inhibitors pepstatin A ($20\text{ }\mu\text{g ml}^{-1}$), leupeptin ($20\text{ }\mu\text{g ml}^{-1}$) and aprotinin ($20\text{ }\mu\text{g ml}^{-1}$), and $1\text{ mM phenylmethylsulphonyl fluoride}$. Insoluble material was removed by centrifugation at $14,000g$ for 10 min at 4°C . The protein content of soluble material was determined by Bio-Rad Dc protein assay. Soluble proteins (approximately $500\text{ }\mu\text{g}$) were incubated overnight with $2.5\text{ }\mu\text{g}$ of anti-AP2 adaptin $\beta 2$ (BD Biosciences). Immune complexes were isolated by addition of $20\text{ }\mu\text{l}$ of protein G-Sepharose beads, followed by incubation for 1–2 h at 4°C . Immunoprecipitates were then washed four times with lysis buffer, resuspended in Laemmli sample buffer, and boiled for 5 min. The samples were subjected to SDS-polyacrylamide gel electrophoresis (PAGE) using a 7% gel, and transferred to a nitrocellulose membrane. Membranes were immunoblotted with a polyclonal antibody to NMDAR2A/2B ($1:500$, Chemicon). The blot was developed using the enhanced chemiluminescence western blot detection system (Amersham). The membrane was then stripped and reprobed with anti-AP2 ($1:3,000$, BD Biosciences).

Miniature EPSC recording

Hippocampal neuron primary cultures were used for electrophysiological recordings 14–18 days after plating as described²⁹ (see also Supplementary Methods).

Received 24 December 2002; accepted 3 February 2003; doi:10.1038/nature01497.

1. Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. The glutamate receptor ion channels. *Pharmacol. Rev.* **51**, 7–61 (1999).
2. Johnson, J. W. & Ascher, P. Glycine potentiates the NMDA response in cultured mouse brain neurons. *Nature* **325**, 529–531 (1987).

3. Kleckner, N. W. & Dingledine, R. Requirement for glycine in activation of NMDA-receptors expressed in *Xenopus* oocytes. *Science* **241**, 835–837 (1988).
4. Carroll, R. C., Lissin, D. V., von Zastrow, M., Nicoll, R. A. & Malenka, R. C. Rapid redistribution of glutamate receptors contributes to long-term depression in hippocampal cultures. *Nature Neurosci.* **2**, 454–460 (1999).
5. Lissin, D. V. *et al.* Activity differentially regulates the surface expression of synaptic AMPA and NMDA glutamate receptors. *Proc. Natl Acad. Sci. USA* **95**, 7097–7102 (1998).
6. Man, H. Y. *et al.* Regulation of AMPA receptor-mediated synaptic transmission by clathrin-dependent receptor internalization. *Neuron* **25**, 649–662 (2000).
7. Ehlers, M. D. Reinsertion or degradation of AMPA receptors determined by activity-dependent endocytic sorting. *Neuron* **28**, 511–525 (2000).
8. Luscher, C. *et al.* Role of AMPA receptor cycling in synaptic transmission and plasticity. *Neuron* **24**, 649–658 (1999).
9. Carroll, R. C. *et al.* Dynamin-dependent endocytosis of ionotropic glutamate receptors. *Proc. Natl Acad. Sci. USA* **96**, 14112–14117 (1999).
10. Kittler, J. T. *et al.* Constitutive endocytosis of GABAA receptors by an association with the adaptin AP2 complex modulates inhibitory synaptic currents in hippocampal neurons. *J. Neurosci.* **20**, 7972–7977 (2000).
11. Roche, K. W. *et al.* Molecular determinants of NMDA receptor internalization. *Nature Neurosci.* **4**, 794–802 (2001).
12. Vissel, B., Krupp, J. J., Heinemann, S. F. & Westbrook, G. L. A use-dependent tyrosine dephosphorylation of NMDA receptors is independent of ion flux. *Nature Neurosci.* **4**, 587–596 (2001).
13. Snyder, E. M. *et al.* Internalization of ionotropic glutamate receptors in response to mGluR activation. *Nature Neurosci.* **4**, 1079–1085 (2001).
14. Shupliakov, O. *et al.* Synaptic vesicle endocytosis impaired by disruption of dynamin-SH3 domain interactions. *Science* **276**, 259–263 (1997).
15. Grabs, D. *et al.* The SH3 domain of amphiphysin binds the proline-rich domain of dynamin at a single site that defines a new SH3 binding consensus sequence. *J. Biol. Chem.* **272**, 13419–13425 (1997).
16. Hansen, S. H., Sandvig, K. & van Deurs, B. Clathrin and HA2 adaptors: effects of potassium depletion, hypertonic medium, and cytosol acidification. *J. Cell Biol.* **121**, 61–72 (1993).
17. Watkins, J. C. & Olverman, H. J. Agonists and antagonists for excitatory amino acid receptors. *Trends Neurosci.* **10**, 265–272 (1987).
18. Mothet, J. P. *et al.* D-serine is an endogenous ligand for the glycine site of the N-methyl-D-aspartate receptor. *Proc. Natl Acad. Sci. USA* **97**, 4926–4931 (2000).
19. Foster, A. C. *et al.* Kynurenic acid analogues with improved affinity and selectivity for the glycine site on the N-methyl-D-aspartate receptor from rat brain. *Mol. Pharmacol.* **41**, 914–922 (1992).
20. Danysz, W. & Parsons, C. G. Glycine and N-methyl-D-aspartate receptors: physiological significance and possible therapeutic applications. *Pharmacol. Rev.* **50**, 597–664 (1998).
21. Slepnev, V. I. & De Camilli, P. Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nature Rev. Neurosci.* **1**, 161–172 (2000).
22. Salituro, F. G. *et al.* 3-(2-Carboxyindol-3-yl)propionic acid-based antagonists of the N-methyl-D-aspartate receptor associated glycine binding site. *J. Med. Chem.* **35**, 1791–1799 (1992).
23. Marks, B. & McMahon, H. T. Calcium triggers calcineurin-dependent synaptic vesicle recycling in mammalian nerve terminals. *Curr. Biol.* **8**, 740–749 (1998).
24. Bekkers, J. M. & Stevens, C. F. NMDA and non-NMDA receptors are co-localized at individual excitatory synapses in cultured rat hippocampus. *Nature* **341**, 230–233 (1989).
25. Mugnaini, M., Dal Forno, G., Corsi, M. & Bunnemann, B. Receptor binding characteristics of the novel NMDA receptor glycine site antagonist [3H]GV150526A in rat cerebral cortical membranes. *Eur. J. Pharmacol.* **391**, 233–241 (2000).
26. Mugnaini, M., Antolini, M., Corsi, M. & van Amsterdam, E. T. [3H]5,7-dichlorokynurenic acid recognizes two binding sites in rat cerebral cortex membranes. *J. Recept. Signal. Transduct. Res.* **18**, 91–112 (1998).
27. Popik, P. *et al.* [3H]1-aminocyclopropanecarboxylic acid, a novel probe for strychnine-insensitive glycine receptors. *Eur. J. Pharmacol.* **291**, 221–227 (1995).
28. Colquhoun, D. Binding, gating, affinity and efficacy: the interpretation of structure-activity relationships for agonists and of the effects of mutating receptors. *Br. J. Pharmacol.* **125**, 924–947 (1998).
29. Huang, Y. *et al.* CACbeta/Pyk2 kinase is a signaling link for induction of long-term potentiation in CA1 hippocampus. *Neuron* **29**, 485–496 (2001).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank J. L. Hicks and Y. Li for technical support, and J. F. MacDonald and L. Y. Wang for comments on the manuscript. This work was supported by the Canadian Institutes of Health Research (CIHR; to M.W.S. and Y.T.W.), the EJLB Foundation (to Y.T.W.), the Ontario Neurotrauma Foundation (ONF; to M.W.S.) and the Nicole Fealdman Memorial Fund (to M.W.S.). M.W.S. is a CIHR Investigator, Y.T.W. is a CIHR Investigator and the holder of the HSF of British Columbia and Yukon Chair in Stroke Research at the University of British Columbia and Vancouver Hospital and Health Sciences Centre, and is an International Scholar of the Howard Hughes Medical Institute. Y.N. is a Ronald Melzack Fellow of CIHR; Y.Q.H. is a Fellow of the ONF and CIHR/Heart and Stroke Foundation of Canada (HSFC); G.A. is a Fellow of Research Training Centre at Hospital for Sick Children; W.J. is a CIHR/HSFC student; and L.V.K. is a CIHR MD/PhD student. We thank J. F. MacDonald and M. Sheng for providing cDNA constructs. Y.T.W. and M.W.S. are joint senior authors.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W.S.
(e-mail: mike.salter@utoronto.ca) or Y.T.W. (e-mail: ytwang@interchange.ubc.ca).

Antibody neutralization and escape by HIV-1

Xiping Wei*, Julie M. Decker*, Shuyi Wang*, Huxiong Hui†, John C. Kappes†‡, Xiaoyun Wu†, Jesus F. Salazar-Gonzalez†, Maria G. Salazar†, J. Michael Kilby†, Michael S. Saag†, Natalia L. Komarova§, Martin A. Nowak§, Beatrice H. Hahn†‡, Peter D. Kwong|| & George M. Shaw*†‡

* Howard Hughes Medical Institute, † Department of Medicine and ‡ Department of Microbiology, University of Alabama at Birmingham, 720 South 20th Street, KAUL 816, Birmingham, Alabama 35294-0024, USA

§ Institute for Advanced Study, Princeton, New Jersey 08540, USA

|| Vaccine Research Center, National Institutes of Health, Bethesda, Maryland 20892, USA

Neutralizing antibodies (Nab) are a principal component of an effective human immune response to many pathogens, yet their role in HIV-1 infection is unclear^{1–6}. To gain a better understanding of this role, we examined plasma from patients with acute HIV infection. Here we report the detection of autologous Nab as early as 52 days after detection of HIV-specific antibodies. The viral inhibitory activity of Nab resulted in complete replacement of neutralization-sensitive virus by successive populations of resistant virus. Escape virus contained mutations in the *env* gene that were unexpectedly sparse, did not map generally to known neutralization epitopes, and involved primarily changes in N-linked glycosylation. This pattern of escape, and the exceptional density of HIV-1 envelope glycosylation generally^{7,8}, led us to postulate an evolving 'glycan shield' mechanism of neutralization escape whereby selected changes in glycan packing prevent Nab binding but not receptor binding. Direct support for this model was obtained by mutational substitution showing that Nab-selected alterations in glycosylation conferred escape from both autologous antibody and epitope-specific monoclonal antibodies. The evolving glycan shield thus represents a new mechanism contributing to HIV-1 persistence in the face of an evolving antibody repertoire.

HIV-1 virions in plasma have an exceedingly short lifespan (<6 hours), as do productively infected lymphocytes (<1.2 days)^{9,10}. As a result, the composition of plasma virus provides a sensitive indicator of biologically relevant selection pressures acting on virus and virus-producing cells^{11–13}. We amplified full-length HIV-1 gp160 envelope genes by polymerase chain reaction (PCR) from serial plasma specimens of patients with acute and early infection, expressed the genes *in trans* with *env*-deficient HIV-1, and tested progeny virus for neutralization susceptibility in a sensitive, single-round viral infectivity assay using JC53BL-13 cells¹³. Figure 1a depicts inhibition of virus entry by sequential autologous plasma specimens from a patient (WEAU), beginning 16 days after onset of symptoms of the acute retroviral syndrome, and four days before detection of HIV-1 antibodies by enzyme-linked immunosorbent assay. No Nab activity was observed in the initial day 16 patient plasma specimen compared with normal donor plasma against an early day 16 virus envelope (16-2). However, increasing titres of Nab were detected in plasma specimens between days 72 and 212, reaching half-maximal and 90% maximal inhibitory concentrations (IC₅₀ and IC₉₀) of 0.0007 ± 0.0002 and 0.0032 ± 0.0023, respectively. Between days 391 and 772, Nab titres to this early virus declined. Nab activity in patient plasma could be eliminated by pre-treatment of plasma with staphylococcal protein G (SPG), reconstituted by IgG eluted from SPG, and virus pseudotyped with the vesicular stomatitis virus G protein (instead of HIV-1 gp160) was not neutralized by patient plasma (data not shown).

The potency of neutralization by WEAU plasma was demonstrated by pre-incubating serial dilutions of plasma from WEAU day 246 (or from a normal uninfected donor) with 1,000, 5,000 or 10,000 infectious units of HIV-1 pseudotyped with either the 16-2 envelope or another early WEAU envelope (1.60), and then tested for entry into 40,000 JC53BL-13 cells. At each multiplicity of infection (0.025–0.25), and for each virus, there was greater than 98% reduction in virus entry at a 1:20 dilution of WEAU plasma, as assessed by both luciferase and β -galactosidase expression. Moreover, there was greater than 99.8% reduction in virus entry at a 1:4 plasma dilution. Normal human plasma at these same dilutions had no effect.

HIV-1 neutralization escape was evaluated using viruses pseudotyped with gp160 envelopes from WEAU plasma specimens 136, 212, 391, 772 and 1,100 days after onset of symptoms. HIV-1 stocks corresponding to these dates were first tested in comparison with virus bearing a day 16 envelope (16-2) for susceptibility to antibody neutralization by contemporaneous plasma samples (Fig. 1b). In each instance, the contemporaneous viruses were less susceptible to neutralization than was the early day 16 virus. These data suggested that changes in the viral envelope had occurred, rendering the successive viruses increasingly resistant to neutralization, or that the host immune system had become tolerant or otherwise unresponsive to evolving variant antigens. Figure 1c shows that both occurred. In the first 391 days of infection, increasing Nab activity was observed against each of the successively emerging virus strains (clones 16-2, 136-14, 212-52 and 391-3), clearly demonstrating *de novo* Nab responses to each new variant. This finding was corroborated using a different day 212 envelope clone (212-55) which was resistant to neutralization by day 136 plasma but increasingly sensitive to neutralization by day 212 and 391 plasmas (Fig. 1d), and this at a time when Nab titres to the early 16-2 virus were declining (Fig. 1a). Later in infection, *de novo* Nab responses could no longer be mounted and pre-existing responses waned (Fig. 1a–c).

To determine the molecular basis of Nab escape, the envelope clones used for Nab analyses were sequenced (Fig. 2). Out of 657 amino acids comprising the gp120/41 ectodomain (positions 30–686), 48 sites were substituted in at least one of the seven clones. Sporadic changes present in two or fewer clones comprised 32 of the mutated sites, and were scattered throughout the gene. Other changes had obviously become fixed over time, including four sets of mutations (positions 30–32; 221; 356; 462–464) corresponding to demonstrated or potential HLA-restricted cytotoxic T lymphocyte (CTL) epitopes in this patient (ref. 12; N. Jones *et al.*, personal communication). Only 1 of 16 fixed changes was in the receptor binding region (position 201), and only 3 mapped to the variable portions of the V1/V2 and V3 loops. Even then, one of these substitutions reverted to the original sequence, and two developed only after more than a year after infection. Instead, most fixed changes were at sites of glycosylation (8 of 16 changes) and on the immunologically silent face (9 of 16 changes)^{14,15}. Substitutions at positions 201 (K to N), 240 (K to T), 268 (N to S), 301 (T to I) and 400–406 (V4), because of their early appearance and persistence, were especially suggestive of selected mutations. Each was associated with changes in potential N-linked glycosylation. On a statistical basis, the predilection for changes in the gp120/41 ectodomain that affected potential N-linked glycosylation sites was non-random. Further evidence for selection on the gp120/41 ectodomain was obtained by sequencing additional WEAU envelope clones corresponding to 16–1,166 days after onset of symptoms (Fig. 3a). Analysis of 105 sequences suggested strong selection pressures at positions 201 and 268 from the outset of infection; V4 beginning by day 136; and position 240 beginning by day 212. Changes at 301 did not appear to be under comparable selection. Fourteen WEAU clones (indicated by asterisks in Fig. 3a) were analysed for autologous Nab sensitivity using day 136 plasma. Clones from day 16

were uniformly sensitive ($IC_{50} < 0.0013$), and later clones uniformly resistant by 10 to 100-fold (Fig. 4a, and data not shown).

To determine if virus neutralization and escape is a general phenomenon in early HIV-1 infection, we examined serial plasma samples from four additional patients who were identified before antibody seroconversion (Table 1). SUMA and BORI, like WEAU, had declined antiretroviral therapy. By 278 and 218 days after onset of symptoms, each had developed IC_{50} Nab titres to autologous early virus of 0.0028 and 0.0010, respectively. For all three subjects, the mean titre of Nab against autologous early viruses (IC_{50} 0.0016 $\pm$ 0.0011) was significantly greater than for autologous late viruses ($IC_{50} > 0.1$; $P < 0.05$) or for the heterologous viruses NL4.3 (0.0094 $\pm$ 0.0047), 16-2 (> 0.1) or Yu2 (> 0.1) ($P < 0.05$ for all). Thus, late viruses from SUMA (day 435), BORI (day 556), and WEAU (day 391), had developed as much as 100-fold (or greater) resistance to autologous Nab. Yet importantly, none of the early viruses was atypically sensitive to neutralization by heterologous

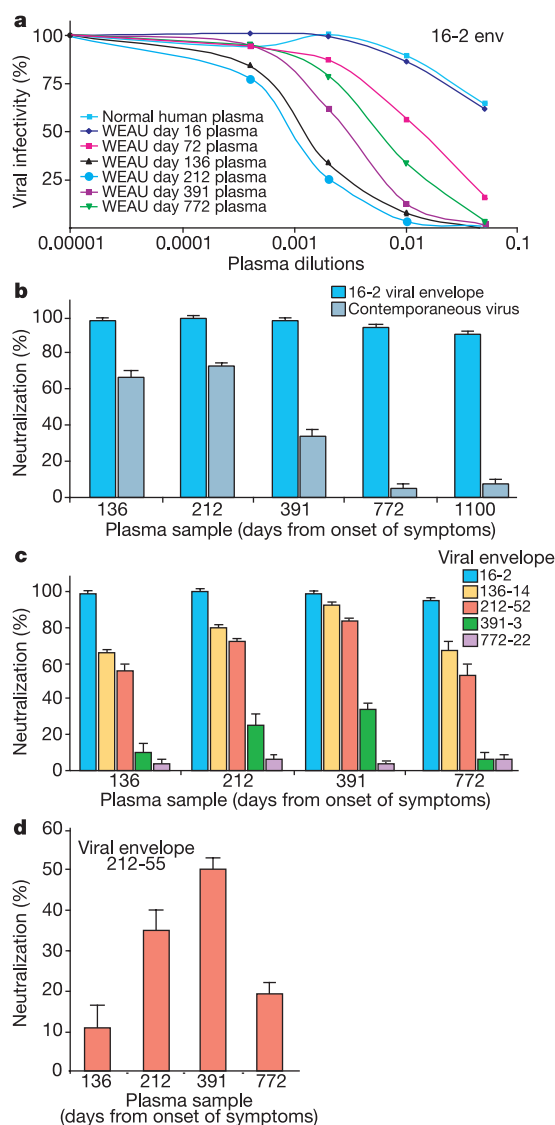


Figure 1 Autologous virus neutralization and escape in subject WEAU. **a**, Titration curves of sequential plasma specimens against virus pseudotyped with HIV-1 envelope from day 16 (16-2). **b**, Neutralization of viruses pseudotyped with early versus contemporaneous envelopes. **c**, Neutralization of viruses with earlier, contemporaneous, or later envelopes. **d**, *de novo* neutralizing antibody response to escape variant 212-55. In **b–d**, subject plasma was diluted 1:50. Virus entry determined by luciferase expression with results expressed as mean $\pm$ 1 s.d.

antibody or by a panel of well characterized monoclonal antibodies, including 17b, F105, b12, 2F5 and 2G12, or by sCD4 (Table 1). Rather, the data indicated that the viral clones from early time-points behaved as neutralization-resistant primary isolates, and that the evolved resistance was not a generalized one but a specific adaptation to the particular Nab in each patient's plasma. Two other patients with acute infection (TRJO and THRO) initiated highly active antiretroviral therapy promptly at diagnosis and before seroconversion; in neither patient were Nab detected. Heterologous Nab activity was examined in 17 additional patients with chronic infection who were not receiving antiretroviral therapy (Table 1). Neutralization was observed in all. NL4.3 was significantly more sensitive to neutralization (IC_{50} 0.0036 ± 0.0030 ; $n = 17$) than was either 16-2 (0.0218 ± 0.0219 ; $P < 0.001$) or Yu2 (0.0613 ± 0.0426 ; $P < 0.001$). There was no correlation between patient CD4 counts or plasma virus loads and heterologous Nab titres.

How can the findings of frequent Nab detection and rapid virus escape at sites distinct from known neutralizing epitopes be reconciled with current models of HIV-1 envelope structure and function? A large body of data shows that Nab epitopes map primarily to the V1/V2 and V3 variable loops on gp120, as well as to the receptor binding site surfaces^{6,15–17}. The pattern of selection that we observed must be interpreted within this context, but it is also necessary to be mindful of the extraordinary density of N-linked glycosylation that segregates to the outer surface of the HIV envelope spike^{15,18}. We thus propose an evolving 'glycan shield' that, within the quaternary confines of the trimer, acts as a steric hindrance preventing Nab from binding neighbouring epitopes, including those at V1/V2, V3, and the receptor binding surfaces. The structure of high mannose glycosylation, which is found predominantly on HIV-1¹⁹, is well-suited to shielding, with a trunk of Asn-(N-acetylglucosamine)-2-mannose extending 15–20 Å from the protein surface before branching into a canopy of 4–8 additional mannoses. Because complete shielding would interfere with receptor binding, HIV must permit holes for functional access. Steric clashes create strong repulsions, and the glycan canopy would only need to partially overlap a potential epitope to prevent access. Shifting the shield to selectively accommodate receptor, but not antibody, alters the accessibility of neutralizing epitopes, and hence neutralization resistance. Such a steric glycan shield would explain the overall neutralization resistance of primary isolates, the notable selection at sites affecting glycosylation or glycan packing, and the lack of selection at actual epitopes of neutralization. Because such a shield would be absolutely dependent on the absence of carbohydrate-reactive antibody, we tested plasma for glycan-reactive antibodies. Previously, it has been shown that the monosaccharide D-mannose can specifically compete with HIV-1 gp120 for binding of the glycan-dependent 2G12 monoclonal antibody²⁰. We found no effect of D-mannose, D-galactose, or N-acetylglucosamine at concentrations of 1, 10 or 100 mM on neutralization by plasmas from any of the 22 subjects or by the b12 monoclonal antibody, while at the same time, there was a dose-dependent D-mannose specific inhibition of virus neutralization by 2G12 (data not shown).

If a glycan shield were responsible for conferring neutralization resistance to HIV-1, then altering sites of N-linked glycosylation by site-directed mutagenesis might lead to Nab resistance. This was done for both WEAU and SUMA. For WEAU, positions 240, 268, 301 and V4 in the Nab-sensitive clone 16-2 were changed to resemble the escape mutant 391-3. Individually, each mutation had only a modest effect, increasing the IC_{50} 1.2–2.6-fold. But together, these substitutions increased the IC_{50} of clone 16-2 by more than 100-fold ($IC_{50} > 0.100$), equivalent to the very resistant clones 391-3 and 391-12 (Fig. 4a). These glycan sites are on average 24 Å from each other. None of the sites is in the receptor binding region or in the V1/V2 or V3 loops. This synergistic behaviour of

such widely spread sites can be explained by a glycan shield that requires cooperative packing over the entire spike to prevent antibody access. We note that elimination of the surface-exposed 301 glycan alone led to a 100-fold decrease in 2G12-mediated neutralization, a finding expected on the basis of its surface exposure and known role in 2G12 binding^{20,21}.

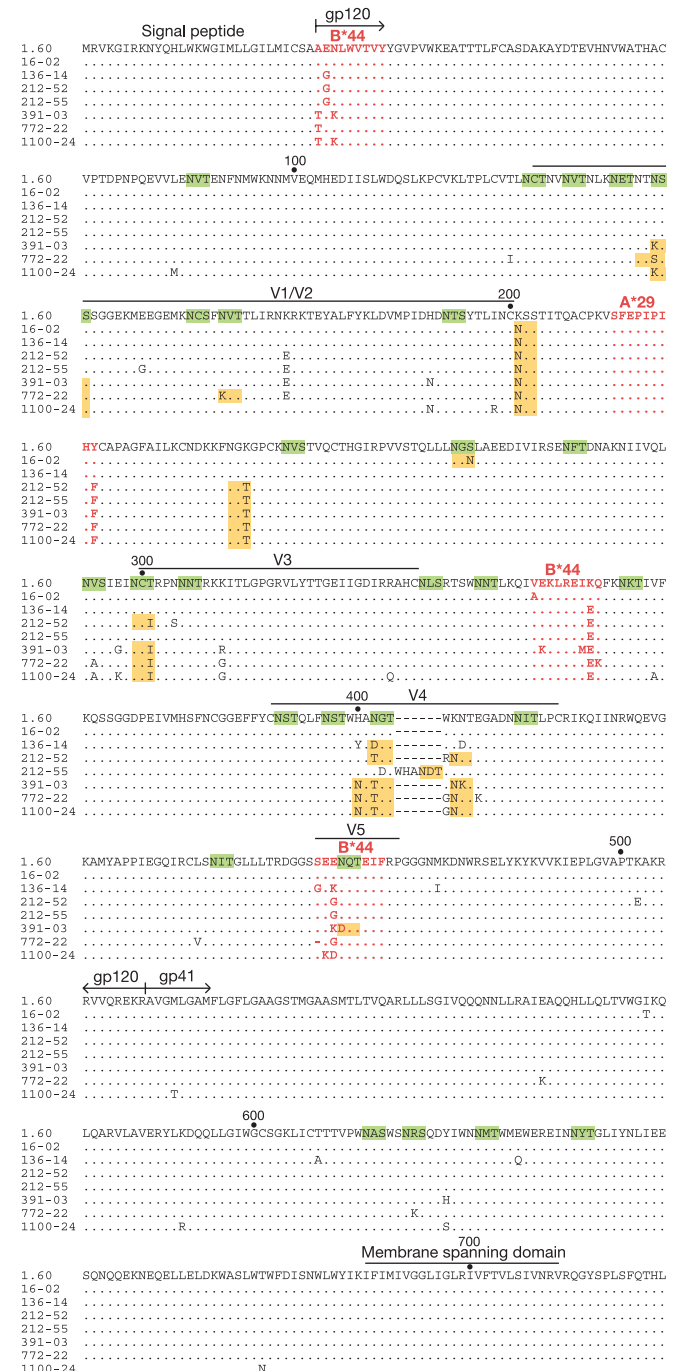


Figure 2 Envelope sequences of sequential WEAU clones spanning days 15–1, 100. Amino acids encompassing the gp120/41 ectodomain are shown. All clones were amplified as intact gp160 genes from uncultured plasma virus except for 1.60, which was obtained as a replication-competent provirus from H9 cells⁷. HLA-restricted CTL epitopes are depicted in red, N-linked glycosylation sites in green, and changes in N-linked glycosylation sites (addition or deletion) in yellow. Amino acid identity (-), insertion/deletion (-) or substitution is indicated.



Figure 3 Partial envelope sequences of sequential plasma-derived clones. **a**, Subject WEAU spanning 1,166 d from acute HIV-1 infection. **b**, Subject SUMA spanning 736 d of follow-up. N-linked glycosylation sites are highlighted in green. Clones used for envelope

expression and pseudotype analysis of neutralization susceptibility are indicated by an asterisk.

Because each N-linked glycan occludes a volume roughly equivalent in size to the V3 loop, the presence or absence of even a single glycan may have a large effect on steric accessibility. Thus, a steric glycan model would predict that neutralization sensitivity could be gained by single changes at sites of glycosylation, but in a context-dependent manner. This was in fact observed. Substitution of a single 16-2 specific amino acid at position 240 (T to K) in clone 212-55 led to a 20-fold increase in neutralization sensitivity, but the same substitution in clone 391-3 had no effect (Fig. 4a).

Although mutations in N-linked glycans can explain much of the Nab resistance in WEAU viruses, alterations in glycan packing resulting from changes in non-glycosylated residues could also contribute. We tested this possibility by probing the K to E substitution at position 356 (Fig. 2), which is close to the inner-domain/outer-domain interface of gp120 on the immunologically silent face¹⁵. As a control, we chose V5 changes because these constitute a surface-exposed structure close to the CD4-receptor binding site. Using clone 391-3 as the backbone, we substituted consensus early sequences at positions 356, V5, V4 and 240, alone or in combination. Single mutations at positions 356 and V4—but not at 240 or V5—led to modest decreases in IC_{50} of 2 to 12-fold. Mutations at 356 and V4 together led to a decrease in IC_{50} of 32-fold. But simultaneous changes at 356, V4 and 240 led to a decrease in IC_{50} of more than 100-fold, to 0.0013 ± 0.0003 , equivalent to early viruses 16-2 and 16-8. Thus, a non-glycan substitution at position 356, along with glycan changes at V4 and 240, was necessary and sufficient to restore complete neutralization sensitivity to an otherwise highly resistant virus.

In SUMA, we also found evidence of selection involving N-linked glycosylation sites in V4 and at positions 235 and 359 on the immunologically silent face (Fig. 3b and Supplementary Information). When each of these mutations was substituted individually into an early day 20 clone (20-12), the pseudotyped viruses became 2 to 5-fold more resistant to autologous neutralization. Together, however, they led to a 35-fold increase in neutralization

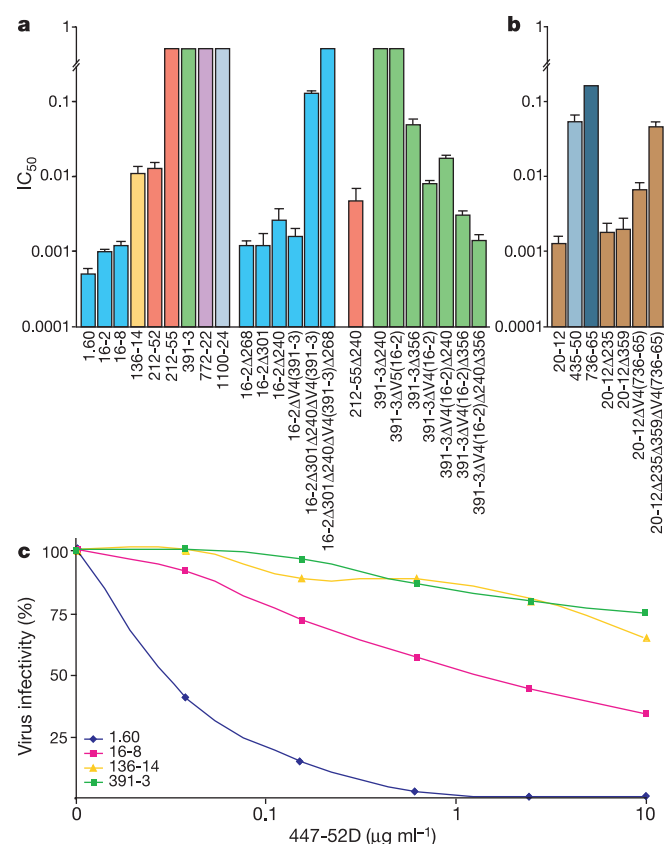


Figure 4 Neutralization of virus pseudotyped with naturally occurring or site-directed mutant envelopes. **a**, WEAU viruses tested with day 136 autologous plasma. **b**, SUMA viruses tested with day 278 autologous plasma. **c**, WEAU viruses tested with monoclonal antibody 447-52D. Virus entry determined by luciferase expression, with results in **a** and **b** expressed as mean $\pm$ 1 s.d. See text for details of mutants.

resistance (Fig. 4b). For BORI, the analysis of Nab escape was complicated by the fact that this individual was infected from the outset with three related but divergent strains of HIV-1 (see Supplementary Information). Despite this heterogeneity, we found evidence for fixation of potential N-linked glycan changes in V1, V2, C3 and V5, as well as additional non-glycan associated changes, all in association with as much as a 100-fold decrease in Nab susceptibility between early and late viruses (Table 1).

Direct evidence in support of a steric glycan shield would include a demonstration that Nab of defined epitope specificity is prevented from binding its cognate epitope by steric hindrance rather than epitope variability. We found three human neutralizing monoclonal antibodies, including 447-52D and 2442, which bind the V3 loop β -hairpin of several HIV-1 strains²² and also neutralize with high potency virus pseudotyped with early *env* clones of WEAU. The most potent of these was 447-52D, which neutralized WEAU 1.60 and 16-8 pseudotyped viruses with IC₅₀ values of 0.06 $\mu\text{g ml}^{-1}$ and 1.63 $\mu\text{g ml}^{-1}$, respectively (Fig. 4c). Clone 1.60 is distinguished from 16-8 and from all other WEAU clones by the absence of an N-linked glycan at position 201 (leading to more open access of the V3 loop to Nab²³) and by an enhanced sensitivity to autologous neutralization (IC₅₀ 0.0004 $\pm$ 0.0001; day 212 plasma). WEAU clones 136-14 and 391-3 are either identical to 1.60 and 16-8 in V3 loop sequence, or in the case of 391-3, differ in a conservative amino acid substitution (K to R) removed from the type II β -turn. Both 136-14 and 391-3 pseudotyped viruses demonstrated greater than 60-fold resistance to 447-52D (IC₅₀ > 100 $\mu\text{g ml}^{-1}$ for each) compared with clone 16-8, and greater than 1,000-fold resistance compared with 1.60. Clones 136-14 and 391-3 were similarly resistant to 2442 (data not shown). These data suggest that changes in envelope conformation and glycan packing, including substitutions at 201, 240, 356 and

V4, can sterically inhibit the accessibility of principal neutralizing epitopes on the virus surface.

The glycan shield model of neutralization escape that we propose builds on a body of evidence implicating glycans in HIV-1 antibody resistance^{8,23–26}. It differs mechanistically from the previously described “silent face”¹⁵ where antibodies are in fact not elicited, presumably because glycosylation resembles ‘self’ closely enough to preclude their elicitation. Rather, the model is distinguished by a repertoire of rapidly evolving, effective, non-glycan-reactive neutralizing antibodies, and a similarly evolving and malleable glycan shield that obstructs neutralizing antibody from binding at neighbouring sites. The steric glycan shield thus represents an extension of a model implicating carbohydrate masking, epitope variation, oligomeric exclusion, and conformational (entropic) masking as complementary mechanisms contributing to antibody resistance²⁷ and virus persistence. The evolving glycan shield model is supported not only by the evidence presented here, but also by analysis of sites of positive selection in HIV-1, which show strong selective pressure on the highly glycosylated silent face and a discordance between neutralization epitopes and sites of positive selection; in contrast, influenza virus haemagglutinin shows extremely high correlation between neutralizing epitopes and sites of positive selection²⁸. Despite enormous variation in HIV-1 viruses generally, the total number of N-linked glycosylation sites on primary isolates remains relatively constant at close to 25 for gp120^{7,8}, suggesting strong selective pressure to maintain this precise level of glycan density, consistent with a global shield. Thus, whereas most viruses evade Nab by mutational variation directly at the epitope of neutralization, we find that HIV-1 uses a fundamentally different mechanism as a means of escape.

Previous studies have examined the early autologous antibody response to HIV-1, and have found evidence for neutralizing

Table 1 Neutralization of HIV-1 by autologous and heterologous plasma, monoclonal antibodies and soluble CD4

		Virus source					
Plasma source	CD4 ⁺	Autologous early*	Autologous late*	Heterologous NL4.3	Heterologous 16-2	Heterologous YU2	
Acute							
WEAU0575	584	0.0010†	>0.1	0.0147	–	>0.1	
SUMA0874	977	0.0028	>0.1	0.0077	>0.1	>0.1	
BORI0637	760	0.0010	>0.1	0.0057	>0.1	>0.1	
‡TRJO4551	629	>0.1	–	>0.1	>0.1	>0.1	
‡THRO4156	538	>0.1	–	>0.1	>0.1	>0.1	
Chronic							
HIDO1099	823	–	–	0.0022	0.0303	>0.1	
BELI1233	627	–	–	0.0045	0.0427	0.0570	
KIMA9001	555	–	–	0.0012	0.0030	0.0016	
WHMA2428	468	–	–	0.0025	0.0739	>0.1	
HAJO0940	428	–	–	0.0018	0.0213	>0.1	
TIMA2466	422	–	–	0.0073	0.0641	0.0473	
SHRO1787	410	–	–	0.0005	0.0118	>0.1	
HYMI2107	349	–	–	0.0042	0.0428	>0.1	
BAMA0037	323	–	–	0.0011	0.0081	0.0026	
SPKE1998	222	–	–	0.0008	0.0008	0.0141	
PUMA1713	208	–	–	0.0014	0.0141	0.0270	
SMST1012	95	–	–	0.0004	0.0034	0.0023	
WATI0855	7	–	–	0.0117	0.0117	>0.1	
YOAL0070	<5	–	–	0.0060	0.0050	0.0105	
RUTH1145	<5	–	–	0.0048	0.0102	0.0804	
LENA1029	<5	–	–	0.0036	0.0063	>0.1	
FARO1042	<5	–	–	0.0066	0.0206	>0.1	
		WEAU		SUMA		BORI	
Monoclonal antibody	16-2	391-3	20-9	435-46	31-15	556-50	NL4.3
17b	>10§	>10	>10	>10	>10	>10	0.4
F105	>10	>10	>10	>10	>10	>10	0.8
b12	0.6	0.4	3.0	0.8	0.8	0.4	0.03
2F5	0.9	1.4	3.3	>10	>10	>10	0.9
2G12	0.1	>10	0.4	0.5	>10	>10	0.5
sCD4	>100	>100	>100	>100	50	>100	0.6
							5.0

activity and (in some cases) virus escape^{1–4,6}. Generally, Nab titres were reported to be low, and slow to develop. In the present study, we show that autologous Nab titres are generally high, rapid in development, sufficiently potent to neutralize >99% of infectious virus *in vitro*, and importantly, of sufficient titres *in vivo* to select for virus escape populations with kinetics comparable to antiretroviral drugs or CTLs (see Supplementary Information)^{11–13,29}. The rapid escape of Nab-resistant virus shown here for subjects WEAU, SUMA and BORI, and in other subjects³⁰, probably occurs in most patients with persistent viraemia. Given the multiplicity of potential pathways and mechanisms of virus escape, it would not be surprising if Nab had little effect on the viral load setpoint of patients with established infection⁵. But in the setting of *de novo* infection resulting from sexual transmission, where the inoculum of infectious virus is low and the efficiency of transmission is exceedingly low (estimated to be less than 1 in 500 exposures), vaccine-elicited Nab could have a far greater impact. □

Methods

Plasma specimens

Blood was collected in acid citrate dextrose, platelet-free plasma prepared by sequential centrifugations at 200g and 1,000g for 10 min each, and 1 ml aliquots stored at -70°C . Before use, plasma was thawed, heat-inactivated at 56°C for 30 min, and clarified by centrifugation at 3,000g for 5 min.

Neutralization assay

Plasma samples were assayed for Nab activity using a modification of a recently described HIV-1 entry assay¹³ that uses the surface adherent HeLa cell-derived JC53BL-13 cell line (NIH AIDS Research and Reference Reagent Program catalogue no. 8129, T2M-bl). JC53BL-13 cells are genetically modified and selected so as to constitutively express CD4⁺, CCR5⁺ and CXCR4⁺. The cells contain integrated luciferase and β -gal genes under tight regulatory control of an HIV-1 LTR, and they are comparable to human peripheral blood mononuclear cells (PBMCs) in susceptibility to infection by R5 and X4 viruses¹³. Pseudotyped virus was prepared in 293T cells and titred by β -gal expression on JC53BL-13 cells, as described¹³. 4×10^4 JC53BL-13 cells were plated in 24-well tissue culture plates (Falcon) and cultured overnight in DMEM supplemented with 10% fetal calf serum (FCS). 1,000 infectious units of pseudotyped virus were combined in a total volume of 125 μl with five-fold dilutions of test plasma beginning at 10% vol/vol in DMEM plus 1% FCS and incubated for 1 h at 37°C . Normal human plasma (NHP) was added to maintain an overall 10% concentration. Virus was then added to JC53BL-13 cells in an equal volume (125 μl) of DMEM plus 1% FCS and 80 $\mu\text{g ml}^{-1}$ DEAE dextran. This brought the concentration of DEAE dextran to 40 $\mu\text{g ml}^{-1}$ and that of human plasma to 5%, which was used as the basis for calculating neutralization titres. After 2 h at 37°C , 400 μl of DMEM plus 10% FCS, test plasma, and NHP were added, keeping the total human plasma concentration at 5%. Cells were incubated at 37°C for 2 d. In some experiments, additional test plasma and NHP were not added back after the two-hour virus-cell co-incubation step; this had a negligible effect on neutralization titres against pseudotyped virus and decreased background effects of NHP. We also adapted the assay to a 96-well format in which 100- μl aliquots of virus-plasma mixture in DMEM containing 5% FCS and 40 $\mu\text{g ml}^{-1}$ DEAE dextran were added to 10^4 JC53BL-13 cells from which media was completely removed. This allowed for scalability and for concentrations of human plasma (test plus normal donor) and FCS to each remain constant at 5% throughout the entire experiment. If replication-competent virus was used, then the entry inhibitor T-20 (enfuvirtide) at 2 $\mu\text{g ml}^{-1}$ was added after 6 h of virus-plasma-cell coinfection to prevent secondary rounds of infection. In all formats, cells were analysed for β -gal or luciferase expression¹³ after two days. Controls included cells exposed to no virus and to virus without test plasma or NHP. Relative infectivity was calculated by dividing the number of luciferase units at each dilution of test plasma by values in wells containing NHP but no test plasma. Data were fitted with a nonlinear function, and IC_{50} , IC_{60} and IC_{90} values calculated by a least-squares regression analysis. All results were replicated in at least three separate experiments.

Cloning and sequencing

Methods for virion RNA preparation, complementary DNA synthesis, PCR amplification and site-directed mutagenesis are given in Supplementary Information.

Statistical analyses

Comparison of IC_{50} values was performed by a Wilcoxon rank sum test and two-tailed Student's *t*-test. Glycosylation frequency determinations were evaluated by exact permutation tests and χ^2 analysis. All calculations were performed in SAS.

Received 6 December 2002; accepted 30 January 2003; doi:10.1038/nature01470.

- Albert, J. *et al.* Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. *AIDS* **4**, 107–112 (1990).
- Pilgrim, A. K. *et al.* Neutralizing antibody responses to human immunodeficiency virus type 1 in primary infection and long-term-nonprogressive infection. *J. Infect. Dis.* **176**, 924–932 (1997).

- Moog, C., Fleury, H. J. A., Pellegrin, I., Kirn, A. & Aubertin, A. M. Autologous and heterologous neutralizing antibody responses following initial seroconversion in human immunodeficiency virus type 1-infected individuals. *J. Virol.* **71**, 3734–3741 (1997).
- Montefiori, D. C., Hill, T. S., Vo, H. T. T., Walker, B. D. & Rosenberg, E. S. Neutralizing antibodies associated with viremia control in a subset of individuals after treatment of acute human immunodeficiency virus type 1 infection. *J. Virol.* **75**, 10200–10207 (2001).
- Poignard, P. *et al.* Neutralizing antibodies have limited effects on the control of established HIV-1 *in vivo*. *Immunity* **10**, 431–438 (1999).
- Parren, P. W. H. I., Moore, J. P., Burton, D. R. & Sattentau, Q. J. The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. *AIDS* **13** (Suppl. A), S137–S162 (1999).
- Kuiken, C. L. *et al.* (eds) *Human Retroviruses and AIDS 2000: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, 2000).
- Myers, G. & Lenroot, R. HIV glycosylation: What does it portend? *AIDS Res. Hum. Retroviruses* **8**, 1459–1460 (1992).
- Ramratnam, B. *et al.* Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *Lancet* **354**, 1782–1785 (1999).
- Markowitz, M. *et al.* A novel antiviral intervention results in a more accurate assessment of HIV-1 replication dynamics and T-cell decay *in vivo*. *J. Virol.* (in the press).
- Wei, X. *et al.* Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117–122 (1995).
- Borrow, P. *et al.* Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature Med.* **3**, 205–211 (1997).
- Wei, X. *et al.* Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* **46**, 1896–1905 (2002).
- Kwong, P. D. *et al.* Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* **8**, 1329–1339 (2000).
- Wyatt, R. *et al.* The antigenic structure of the human immunodeficiency virus gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
- Wyatt, R. & Sodroski, J. The HIV-1 envelope glycoproteins: fusogens, antigens and immunogens. *Science* **280**, 1884–1888 (1998).
- Poignard, P., Saphire, E. O., Parren, P. W. H. I. & Burton, D. R. GP120: Biologic aspects of structural features. *Annu. Rev. Immunol.* **19**, 253–274 (2001).
- Kwong, P. D., Wyatt, R., Sattentau, Q. J., Sodroski, J. & Hendrickson, W. A. Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *J. Virol.* **74**, 1961–1972 (2000).
- Mizuuchi, T. *et al.* Diversity of oligosaccharide structures on the envelope glycoprotein gp120 of human immunodeficiency virus 1 from the lymphoblastoid cell line H9. Presence of complex-type oligosaccharides with bisecting N-acetylglucosamine residues. *J. Biol. Chem.* **265**, 8519–8524 (1990).
- Scanlan, C. N. *et al.* The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of $\alpha \rightarrow 2$ mannose residues on the outer face of gp120. *J. Virol.* **76**, 7306–7321 (2002).
- Sanders, R. W. *et al.* The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J. Virol.* **76**, 7293–7305 (2002).
- Gorny, M. K. *et al.* Human monoclonal antibodies specific for conformation-sensitive epitopes of V3 neutralize human immunodeficiency virus type 1 primary isolates from various clades. *J. Virol.* **76**, 9035–9045 (2002).
- Kolchinsky, P., Kiprilov, E. & Sodroski, J. Increased neutralization sensitivity of CD4-independent human immunodeficiency virus variants. *J. Virol.* **75**, 2041–2050 (2001).
- Chackerian, B., Rudensey, L. M. & Overbaugh, J. Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host alter recognition by neutralizing antibodies. *J. Virol.* **71**, 7719–7727 (1997).
- Back, N. K. *et al.* An N-glycan within the human immunodeficiency virus type 1 gp120 V3 loop affects virus neutralization. *Virology* **199**, 431–438 (1994).
- Reitter, J. N., Means, R. E. & Desrosiers, R. C. A role for carbohydrates in immune evasion in AIDS. *Nature Med.* **4**, 679–684 (1998).
- Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678–682 (2002).
- Stewart, J. J., Watts, P. & Litwin, S. *Profile of a Quasispecies: The Selected Mutational Landscape of HIV and Influenza* (Vrije Univ., Amsterdam, 1999).
- Novak, M. A. & May, R. M. *Viral Dynamics* (Oxford Univ. Press, Oxford, 2000).
- Richman, D., Wrinn, T., Little, S. & Petropoulos, C. Rapid evolution of the neutralizing antibody response to human immunodeficiency virus type 1 infection. *Proc. Natl Acad. Sci. USA* (in the press).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank S. Zolla-Pazner for HIV-1 V3 specific monoclonal antibodies; D. Burton, B. Korber, J. Mascola, D. Montefiore, J. Moore, L. Shapiro and R. Wyatt for discussions; S. Meleth for assistance with statistical analyses; and W. Abbott for artwork and technical assistance. This work was supported by the Acute Infection Early Disease Research Program of the National Institutes of Allergy and Infectious Diseases, the UAB Center for AIDS Research, the UAB General Clinical Research Center, and the National Institutes of Health. We also thank Bristol-Myers Squibb, GlaxoSmithKline, Agouron and Merck for continuing support of acute HIV-1 infection studies.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.M.S. (e-mail: gshaw@uab.edu).

Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer

D. Neil Watkins*, David M. Berman†‡, Scott G. Burkholder*, Baolin Wang‡, Philip A. Beachy‡ & Stephen B. Baylin*

* Sidney Kimmel Comprehensive Cancer Center, † Department of Pathology, ‡ Department of Molecular Biology and Genetics, and Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA

Embryonic signalling pathways regulate progenitor cell fates in mammalian epithelial development and cancer^{1,2}. Prompted by the requirement for sonic hedgehog (Shh) signalling in lung development^{3,4}, we investigated a role for this pathway in regeneration and carcinogenesis of airway epithelium. Here we demonstrate extensive activation of the hedgehog (Hh) pathway within the airway epithelium during repair of acute airway injury. This mode of Hh signalling is characterized by the elaboration and reception of the Shh signal within the epithelial compartment, and immediately precedes neuroendocrine differentiation. We reveal a similar pattern of Hh signalling in airway development during normal differentiation of pulmonary neuroendocrine precursor cells, and in a subset of small-cell lung cancer (SCLC), a highly aggressive and frequently lethal human tumour with primitive neuroendocrine features. These tumours maintain their malignant phenotype *in vitro* and *in vivo* through ligand-dependent Hh pathway activation. We propose that some

types of SCLC might recapitulate a critical, Hh-regulated event in airway epithelial differentiation. This requirement for Hh pathway activation identifies a common lethal malignancy that may respond to pharmacological blockade of the Hh signalling pathway.

Sonic hedgehog (Shh), a mammalian hedgehog (Hh) pathway ligand, mediates epithelial–mesenchymal interactions in lung development by signalling to adjacent lung mesenchyme, as indicated by expression of the Hh receptor and pathway target *Patched* (*Ptch*)⁵. Loss of Shh function results in severe lung defects associated with failure of branching morphogenesis^{3,4}. As developmental pathways regulate progenitor cell fates and differentiation in some regenerating mammalian epithelia^{1,2}, we hypothesized that Hh signalling might be important in airway epithelial repair.

In contrast to the skin and colon, adult airway epithelium rarely proliferates unless injured⁶. To uncover a role for Hh signalling in this process, we studied a mouse model of acute airway repair in which Clara cells, specialized airway epithelial cells predominant in distal conducting airways, are depleted within 24 h of systemic naphthalene administration⁶. Activation of a putative airway progenitor results in epithelial regeneration within three days, with increased numbers of airway neuroendocrine cells—a normally rare cell type implicated in the regulation of airway epithelial growth and development^{6,7}. In regenerating airways, we observed marked expression of both Shh ligand and Gli1, a transcriptional target of Hh signalling⁸, in the epithelial compartment 72 h after naphthalene injury (Fig. 1a). By day 4, Gli1 was not observed in nascent airway epithelial cells expressing calcitonin gene-related peptide (CGRP), a marker of neuroendocrine differentiation (Fig. 1a, b). These data show that acute airway epithelial regeneration results in widespread activation of airway intraepithelial Hh signalling, which

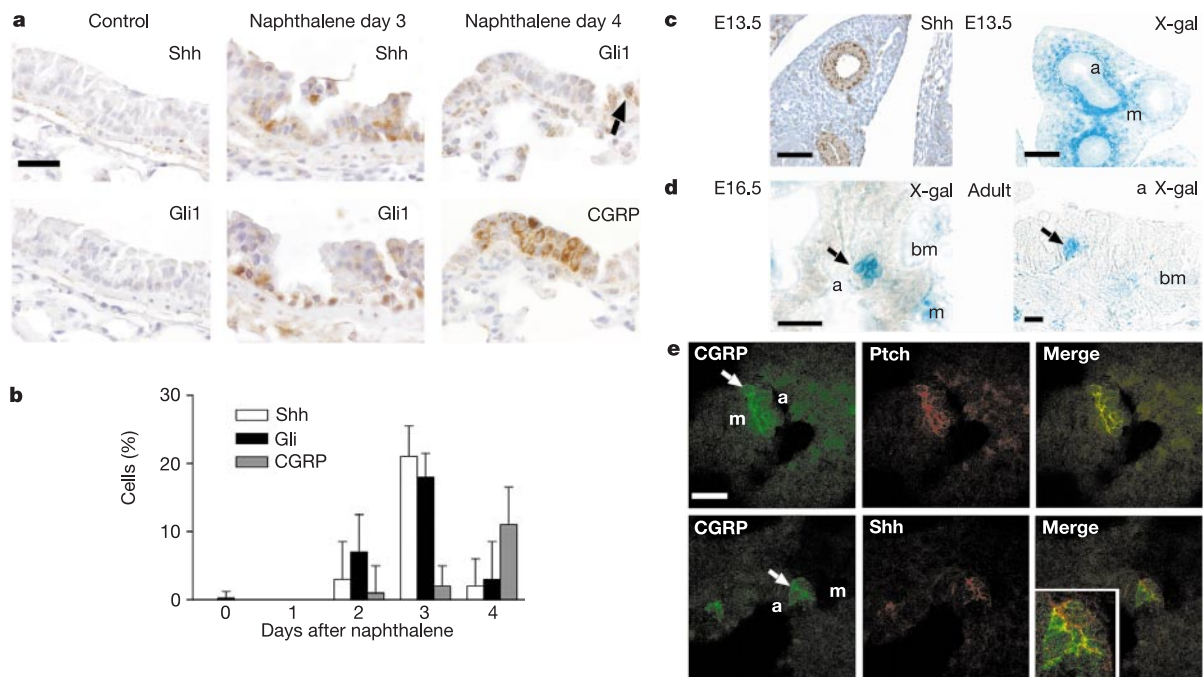


Figure 1 Hedgehog signalling in airway repair and development. **a**, Immunohistochemical detection of Shh and Gli1 in adult mouse airways is negative in normal airways (left panels), but positive for both Shh and Gli1 in serial sections 3 days after naphthalene injury (middle panels). By 4 days after naphthalene treatment (right panels), Gli1-positive cells are reduced in number (arrow). Serial sections demonstrate that nascent CGRP-positive cells do not express stained Gli1. Scale bar, 50 μ m. **b**, Quantitative analysis of bronchial epithelial staining in **a** ($n = 4$, mean $\pm$ s.e.m.). **c**, Shh signalling in E13.5 lungs. Shh immunostaining in embryonic airway epithelium is shown in the left panel. The

right panel shows X-gal staining of lungs obtained from E13.5 *Ptch-LacZ* mouse embryos, demonstrating intense mesenchymal staining. Scale bar, 25 μ m. **d**, Clusters of LacZ-positive cells (arrows) are seen in the airway epithelium of E16.5 (left panel) and adult (right panel) mice. Scale bar, 25 μ m. a, airway; m, mesenchyme; bm, basement membrane. **e**, Confocal immunofluorescence detection of Hh signalling in lung development. The top row demonstrates expression of both CGRP and *Ptch* in an E16.5 airway (arrow), similar to that shown in **d**. The bottom row shows expression of CGRP (arrow) adjacent to Shh-expressing epithelial cells (see high-magnification inset). Scale bar, 25 μ m.

immediately precedes neuroendocrine differentiation.

Embryonic lung epithelial cells express Shh, which is thought to signal to adjacent lung mesenchyme to regulate branching morphogenesis^{3–5}. In light of this, our detection of Shh and Gli1 within the epithelial compartment during airway epithelial regeneration was unexpected. To determine whether such intraepithelial signalling occurred in embryonic lung development, we studied mice in which one copy of *Ptch* is replaced in-frame with the β -galactosidase (β -gal) gene by homologous recombination⁹. As *Ptch* is a transcriptional target of the Gli proteins, expression of β -gal indicates activation of the Hh pathway^{9,10}. Early gestation (embryonic day (E)13.5) embryos showed expression of Shh protein in the primitive lung endoderm, and intense β -gal expression in the adjacent mesenchyme (Fig. 1c). By contrast, later lung development (E16.5) was characterized by clusters of β -gal-expressing cells in the developing airway epithelium (Fig. 1d). Small numbers of cells expressing β -gal persist in the basal layer of the adult bronchial epithelium (Fig. 1d). Similar clusters of epithelial cells expressing the neuroendocrine marker CGRP and *Ptch* were observed by confocal immunofluorescence in E16.5 airways, immediately adjacent to cells expressing Shh (Fig. 1e). These data suggest that during normal development, neuroendocrine precursors within the airway epithelial compartment respond to a Shh signal elaborated by adjacent airway epithelial cells.

SCLC is an aggressive, highly lethal malignancy with primitive neuroendocrine features¹¹. As aberrant reactivation of developmental pathways may have a role in cancer growth^{1,2}, we wondered whether the epithelial Hh signalling we had observed in airway embryogenesis and repair might persist in SCLC. Analysis of SCLC tissue showed that five out of ten tumours expressed both Shh and Gli1 (Fig. 2a; see also Supplementary panel a). Out of 40 non-SCLC (NSCLC) tumours, nine demonstrated Shh expression and of

these, four demonstrated co-expression of Gli1 (Fig. 2a; see also Supplementary panel a). These data provide indirect evidence of persistent activation of Hh signalling in lung cancer, predominantly in SCLC. These findings were confirmed by analysis of human lung cancer cell lines. Notably, all seven SCLC and seven NSCLC cell lines expressed Shh protein (Fig. 2b). Out of five breast and eight colon cancer cell lines examined, only one (CACO2) expressed Shh protein, and none expressed Gli1 protein, as shown by western blot analysis (data not shown). Importantly, expression of both Shh and Gli1 proteins was observed in five out of seven SCLC lines, and this correlated with increased expression of *Ptch* messenger RNA (Fig. 2b). In contrast, NSCLC lines expressed Shh and low levels of *Ptch*, but not Gli1. These data are summarized in Supplementary panel b.

To determine how Hh signalling might function in these tumours, we co-cultured cancer cells with Shh-LIGHT2 cells, a fibroblast reporter cell line that responds to exogenous Shh by activation of an integrated Gli-responsive luciferase reporter¹⁰. Some NSCLC cells that express Shh are capable of heterologous cell signalling to the reporter cell line (Fig. 2c), suggesting that NSCLC retains the Shh export properties of primitive lung endodermal cells that signal to adjacent mesenchymal cells in early development. By contrast, the SCLC cells we studied display a marked reduction in this ability to signal to adjacent cells. These data demonstrate that distinct types of lung cancer cells recapitulate different aspects of Shh signalling seen in lung development and repair.

We next addressed the mechanism of Hh pathway activation in SCLC. Dual-label immunostaining for Shh and Gli1 in SCLC nude mouse xenografts demonstrated Shh-expressing cells adjacent to Gli1-expressing cells (Fig. 2d). These data suggest juxtacrine Hh pathway activation in SCLC markedly similar to that observed in

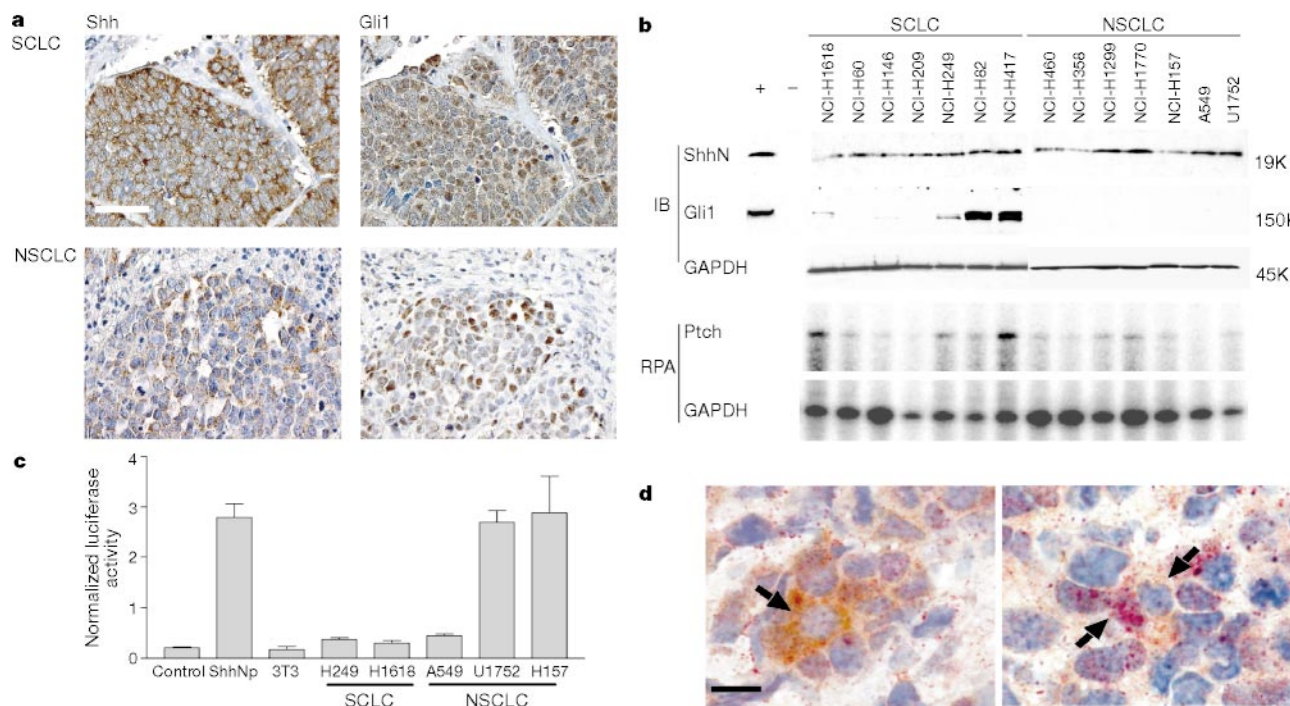


Figure 2 Hh signalling in lung cancer. **a**, Examples of Shh and Gli1 immunostaining in human lung cancer tissue. Note the widespread co-expression of Shh and Gli1 in SCLC, which is reduced in the NSCLC example. **b**, Expression of Hh signalling components in lung cancer cell lines. The top panel shows immunoblotting (IB) data for expression of Shh, Gli1 and GAPDH. The bottom panel demonstrates *Ptch* mRNA expression in the same cell lines detected by RNase protection assay (RPA). The markers along the right indicate relative molecular mass. **c**, Induction of *Gli-luciferase* activity in Shh-LIGHT2 reporter cells

co-cultured with purified Shh-Np or the cell lines indicated on the x axis. Luciferase activity is normalized to a Renilla luciferase internal control ($n = 6$, mean $\pm$ s.e.m.). **d**, Shh and Gli1 expression in NCI-H249 SCLC xenograft cells detected by dual-label immunohistochemistry. Brown, Shh; red, Gli1. The left panel shows a tumour cell expressing Shh alone (arrow); the right panel shows a Shh-expressing tumour cell (top arrow) and an adjacent Gli1-expressing tumour cell (bottom arrow).

airway development and repair. Next, we asked whether ligand-driven Hh pathway activation promotes growth of SCLC. Inhibition of Shh ligand activity in NCI-H249 and NCI-H1618 SCLC cells with the 5E1 Shh-N monoclonal antibody¹² resulted in growth inhibition (Fig. 3a). Although NCI-H157 NSCLC cells express Shh, they do not express Gli1 protein, and are not affected by 5E1 treatment (Fig. 3a). These data demonstrate that growth of SCLC cells *in vitro* is dependent on ligand-mediated activation of the Hh pathway, and suggest the presence of a normal Ptch receptor, confirmed by sequencing of *Ptch* complementary DNA in both NCI-H249 and NCI-H1618 SCLC cells generated by reverse transcription-polymerase chain reaction (RT-PCR) (data not shown).

The *Veratrum* alkaloid cyclopamine specifically inhibits the Hh pathway^{10,13,14} through interaction with the Hh signalling protein smoothened^{15,16}. Moreover, cyclopamine blocks the oncogenic effects of mutations of *Ptch* in fibroblasts¹⁰, and inhibits the malignant growth of medulloblastoma cells lacking *Ptch* function¹⁷. Treatment of NCI-H249 SCLC cells with cyclopamine, or a more potent analogue KAAD-cyclopamine¹⁰, resulted in significant growth inhibition, whereas tomatidine, a closely related compound that lacks the capacity to inhibit Hh signalling, had no effect (Fig. 3b). The effects of cyclopamine and KAAD-cyclopamine on the growth of SCLC reflect their relative potency in silencing Hh pathway activation *in vitro*¹⁰. None of KAAD-cyclopamine, cyclopamine or tomatidine was able to affect growth of NCI-H157 NSCLC cells (Fig. 3c). The growth-inhibitory effect of cyclopamine, if due to Hh pathway blockade, should be bypassed by constitutive overexpression of the Hedgehog pathway effector Gli1 (ref. 17). We indeed observed that stable expression of a Flag-tagged Gli1 protein¹⁸ protected NCI-H249 SCLC cells from growth inhibition by cyclopamine, whereas a Gli1 mutant lacking the zinc finger DNA-binding domain had no effect (Fig. 3d). Treatment of nine cancer cell lines with cyclopamine at concentrations up to 10 μ M demonstrated growth inhibition only in SCLC cells that expressed both Shh and its transcriptional effector Gli1 (Supplementary panel b). These data show that cyclopamine induces growth inhibition in SCLC cells expressing both Shh and Gli1 by specific inhibition of the Hh pathway.

We next investigated the relationship between Hh pathway blockade by cyclopamine and growth arrest in SCLC. Unsynchronized NCI-H249 SCLC cells treated with 5 μ M cyclopamine for 72 h demonstrated arrest of the cell cycle in Go/G1 (Fig. 3e) and apoptosis indicated by an increase in cleaved PARP (Fig. 3f). Analysis of *Ptch* mRNA expression revealed downregulation in response to cyclopamine treatment (Fig. 3g). These results indicate silencing of Hh pathway activation at concentrations of cyclopamine that induce both growth arrest and apoptosis. We next investigated the possibility that SCLC cells might express transcripts indicative of a progenitor cell phenotype. We detected expression of BMP4, a morphogen and putative target of Hh expressed in lung epithelial embryogenesis¹⁹, and nestin, an intermediate filament characteristic of neural stem cells in medulloblastoma¹⁷ (Fig. 3h). Treatment of NCI-H249 SCLC cells with cyclopamine for 48 h inhibited expression of both these genes (Fig. 3h), as well as the expression of human ASH-1, a transcription factor required for pulmonary neuroendocrine differentiation²⁰. These changes in gene expression suggest that Hh signalling maintains a progenitor cell fate in SCLC.

Pathological activation of Hh signalling is associated with medulloblastoma, a malignant brain tumour thought to arise from the granule cells of the cerebellum^{9,21}. Maintenance of abnormal progenitor-like fates through continued Hh pathway activation is essential for malignant growth of these tumours *in vivo*¹⁷. We wondered whether SCLC cells were similarly dependent on Hh signalling for their malignant behaviour. NCI-H249 SCLC cells treated with cyclopamine showed reduced soft agar clonogenicity—an *in vitro* assay of tumorigenicity (Fig. 4a, b). This effect was

reversed in cells overexpressing the Hh pathway transcriptional effector Gli1 (Fig. 4a, b). We next tested the ability of systemic cyclopamine treatment to inhibit the growth of SCLC xenografts in nude mice. Mice bearing xenografts were treated subcutaneously

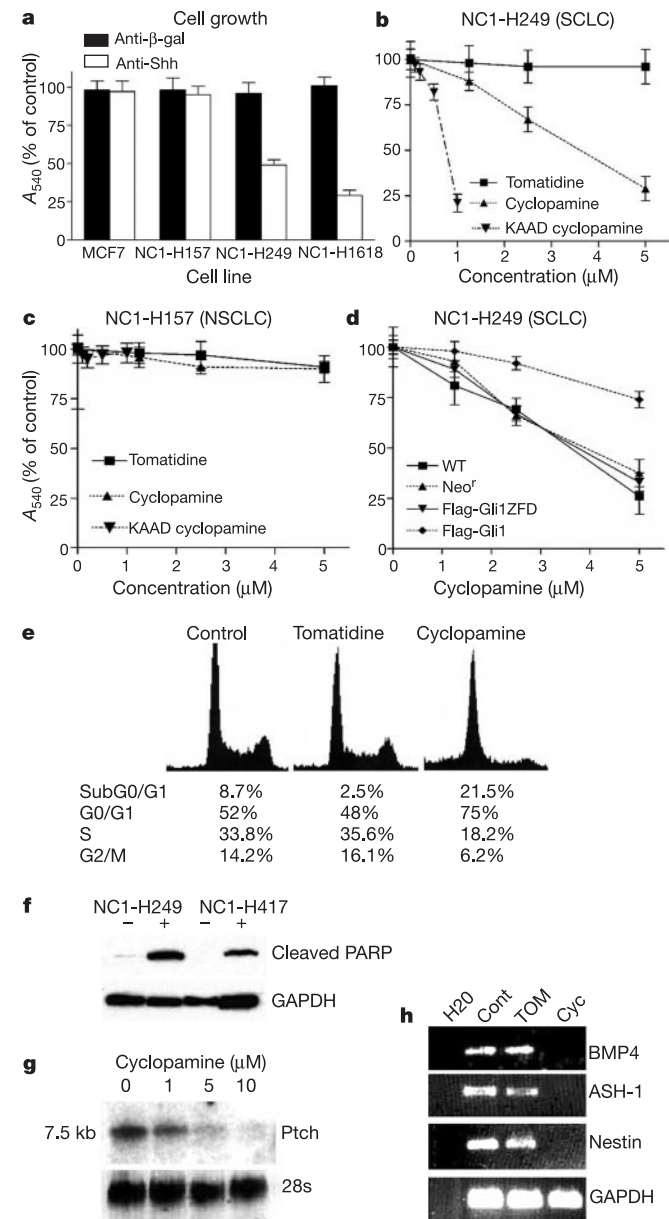


Figure 3 Hh pathway activation is essential for the growth of SCLC. **a**, Growth of cancer cell lines treated with monoclonal antibodies against β -galactosidase (β -gal) as a control, or Shh for 4 days. **b**, NCI-H249 SCLC cell growth after 5 days, treated with tomatidine, cyclopamine or KAAD cyclopamine at the indicated concentrations. **c**, Identical experiment performed in NCI-H157 NSCLC cells. **d**, Response of stably transfected NCI-H249 SCLC cells to treatment with cyclopamine when expressing neomycin resistance (Neo^r), a mutant Gli1 lacking the zinc finger domain (Flag-Gli1ZFD), Gli1 (Flag-Gli1), and wild-type untransfected (WT) cells. Cell viability was measured by MTT assay, detected at an absorbance at 540 nm (A_{540}) ($n = 6$) and expressed as a percentage of control $\pm$ s.e.m. **e**, Cell cycle analysis in NCI-H249 cells treated with tomatidine or cyclopamine (5 μ M). Percentages in each phase of the cell cycle are shown below and are shown as the mean of three experiments. **f**, Cleaved PARP expression in NCI-H249 and NCI-H417 SCLC cells treated with tomatidine (–) or cyclopamine (+) (5 μ M). **g**, *Ptch* mRNA expression in NCI-H249 SCLC cells detected by northern blot analysis after treatment with cyclopamine. 28s RNA stained with ethidium bromide is shown as a loading control. **h**, RT-PCR analysis of transcripts in NCI-H249 SCLC cells. Cont, control; Tom, tomatidine treated; Cyc, cyclopamine treated.

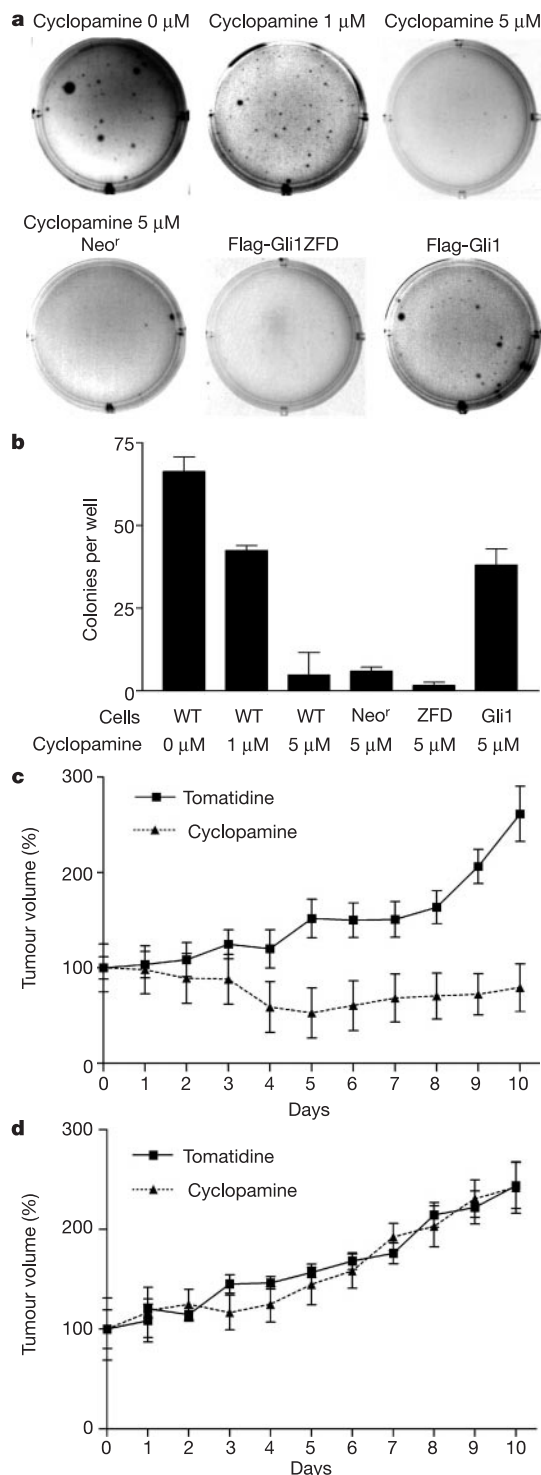


Figure 4 Cyclopamine inhibits SCLC tumorigenicity. **a**, The top panel shows soft agar growth of NCI-H249 SCLC cells treated with cyclopamine. Plates were stained with ethidium bromide. The bottom panel shows colony formation of NCI-H249 SCLC cells treated with cyclopamine (5 μM) and stably transfected with neomycin resistance (Neo^r), mutant Gli1 (Flag-Gli1ZFD) or Gli1 (Flag-Gli1). **b**, Quantitative analysis of the experiment described in **a**. Data are shown as mean colonies per well ± s.e.m. ($n = 6$). **c**, Growth of NCI-H249 nude mouse subcutaneous xenografts in animals treated with tomatidine or cyclopamine for 10 days. **d**, Identical experiment to that shown in **c** except that A549 NSCLC cells were used. Data are shown as mean tumour volume ± s.e.m. as a percentage of tumour volume at day 0 ($n = 7$).

with 25 mg⁻¹ kg⁻¹ day⁻¹ cyclopamine as described¹⁷. Growth inhibition was observed in three SCLC lines: NCI-H249 (Fig. 4c), as well as NCI-H417 and NCI-H1618 (data not shown). No effect was observed in A549 NSCLC cells (Fig. 4d) nor in HCT-116 colon cancer xenografts (data not shown). These data are summarized in Supplementary panel b, and show that Hh signalling is required for the growth *in vivo* of SCLC cells that express both Shh and Gli1.

We have shown that Hh signalling in airway epithelium is not limited to epithelial-mesenchymal interactions, but can be contained within the airway epithelial compartment during embryonic neuroendocrine differentiation and airway repair. Taking evidence that links Hh signalling to cerebellar progenitor cell differentiation into consideration²¹⁻²³, we propose a similar role for this pathway in the regulation of airway progenitor cell fates, which may be specified immediately before the divergence of neuroendocrine and non-neuroendocrine lineages. The dependency of SCLC cells on Hh pathway activation is also notable in that it relies on the presence of Shh ligand, it occurs in the absence of mutations in *Ptch*, and recapitulates juxtacrine Hh signalling seen in development and airway repair. SCLC may represent a malignancy arising from an airway epithelial progenitor that retains both Hh signalling and primitive features of pulmonary neuroendocrine differentiation. The vulnerability of SCLC to Hh pathway blockade may represent a new therapeutic approach to a disease with a poor prognosis²⁴. □

Methods

Detection of β-gal expression

Ptch-LacZ mice were maintained and genotyped as described⁹. 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) staining in microdissected mouse lungs was performed overnight as described²⁵, followed by post-fixation in formalin, paraffin embedding and sectioning. We used wild-type littermates as negative controls.

Immunohistochemistry

Single-colour DAB-immunoperoxidase staining was performed using a modification of the DAKO CSA system. Detailed protocols are available on request. Antibodies were from Santa Cruz Biotechnologies: Shh (N-19; sc-1194); Gli1 (N-16; sc-6153); CGRP (N-20; sc-8856). Shh, Ptch and Gli1 staining was optimized on paraffin sections from Shh wild-type and knockout embryos. Gli1 staining was further confirmed in Flag-Gli1-overexpressing Cos-7 cells by immunofluorescence. Peptide competition ablated staining in tumour samples and embryos. Dual-colour immunohistochemistry was performed using the DAKO Envision system. Dual-colour immunofluorescence was performed on fresh-frozen sections fixed in paraformaldehyde using Molecular Probes Alexa secondary antibodies.

Western immunoblot

Whole-cell lysates were sonicated in 2% SDS/50 mM TrisHCl, pH 8. Western blot using rabbit polyclonal antibodies for Shh-N were performed as described²⁶. A rabbit polyclonal antibody to Gli1 was developed as described²⁷ using a glutathione S-transferase fusion protein containing amino acid residues 216–271 of human Gli1. Anti-cleaved PARP was obtained from Promega.

RNAse protection assay and northern blot analysis

RNAse protection assay (RPA) was performed as described²⁸ using a *Ptch*-specific antisense RNA probe corresponding to bases 1338–1788 of the human patched-1 cDNA (GI:1335863) generated by RT-PCR and subcloned into pCR-TOPOII (Stratagene). Northern blotting of 10 μg total RNA was performed as described²⁹, and probed with a *Ptch*-specific cDNA probe obtained from the same construct.

RT-PCR

Total cellular RNA was treated with DNase, reverse transcribed, and amplified for 31 cycles at an annealing temperature of 55 °C. Primers used were BMP4(+), 5'-CTTTACCGGC TTCAGTCTGGG-3'; BMP4(-), 5'-CCCAATTCCTCACTCCCTTGAG-3'; GAPDH(+), 5'-ATCTTCAGGAGCGAGATCCC-3'; GAPDH(-), 5'-CGTTCGGCTCAGGGATGA CCT-3'; ASH-1(+), 5'-CGCATGGAAGCTCTGCCAAG-3'; ASH-1(-), 5'-TGACC AACTTGACGCGGTTGC-3'; nestin(+), 5'-CTCTGGGAGAGGATCAAG-3'; nestin(-), 5'-CCTTTGTGAGAGGTCTCAGTG-3'.

Shh-LIGHT2 reporter assay

Superconfluent reporter cells were cultured as described¹⁰, then co-cultured in low serum conditions in the presence of 1×10^5 cells per well of the cell line of interest or purified Shh-Np¹⁰. Luciferase and Renilla luciferase assays were performed using the Promega Dual Luciferase Reporter Assay system.

Cell culture experiments

Cell lines were obtained from American Type Culture Collection (ATCC). Shh inhibitor experiments were performed in 0.5% calf serum. Cyclopamine was obtained from

Toronto Research Chemicals. Both were dissolved as $\times 1,000$ stocks in DMSO medium. Flag-tagged Gli1 vectors were obtained from the Joyner laboratory¹⁸. To generate NCI-H249 SCLC cells overexpressing each of the Gli vectors, mass cultures were stably co-transfected using lipofectamine (Invitrogen) with the Flag-Gli vector of interest, and pcDNA3.1 (Stratagene) to confer neomycin resistance. 5E1 anti-ShhN monoclonal antibody was used at a concentration of $10 \mu\text{g ml}^{-1}$ as described¹². Soft agar assays were performed as described¹⁰. Cells were seeded into six-well plates at a density of 20,000 cells per well in agar containing 2% calf serum. MTT assays were performed as described²⁸.

Nude mouse xenografts

Tumour cell lines were injected subcutaneously at 1×10^7 cells per mouse and allowed to grow to a maximum diameter of 5 mm. Cyclopamine was administered as described¹⁷. Tumours were measured daily and the tumour volume calculated as described³⁰.

Received 1 November 2002; accepted 12 February 2003; doi:10.1038/nature01493.
Published online 5 March 2003.

1. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
2. Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001).
3. Piccoli, C. V., Lewis, P. M. & McMahon, A. P. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr. Biol.* **8**, 1083–1086 (1998).
4. Litlington, Y., Lei, L., Westphal, H. & Chiang, C. Sonic hedgehog is essential to foregut development. *Nature Genet.* **20**, 58–61 (1998).
5. Bellusci, S. *et al.* Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. *Development* **124**, 53–63 (1997).
6. Reynolds, S. D., Giangreco, A., Power, J. H. & Stripp, B. R. Neuroepithelial bodies of pulmonary airways serve as a reservoir of progenitor cells capable of epithelial regeneration. *Am. J. Pathol.* **156**, 269–278 (2000).
7. Peake, J. L., Reynolds, S. D., Stripp, B. R., Stephens, K. E. & Pinkerton, K. E. Alteration of pulmonary neuroendocrine cells during epithelial repair of naphthalene-induced airway injury. *Am. J. Pathol.* **156**, 279–286 (2000).
8. Lee, J., Platt, K. A., Censullo, P. & Ruiz i Altaba, A. Gli1 is a target of Sonic hedgehog that induces ventral neural tube development. *Development* **124**, 2537–2552 (1997).
9. Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
10. Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).
11. Zochbauer-Muller, S., Gazdar, A. F. & Minna, J. D. Molecular pathogenesis of lung cancer. *Annu. Rev. Physiol.* **64**, 681–708 (2002).
12. Ericson, J., Morton, S., Kawakami, A., Roelink, H. & Jessell, T. M. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* **87**, 661–673 (1996).
13. Incardona, J. P., Gaffield, W., Kapur, R. P. & Roelink, H. The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. *Development* **125**, 3553–3562 (1998).
14. Cooper, M. K., Porter, J. A., Young, K. E. & Beachy, P. A. Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science* **280**, 1603–1607 (1998).
15. Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16**, 2743–2748 (2002).
16. Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothened activity. *Proc. Natl Acad. Sci. USA* **99**, 14071–14076 (2002).
17. Berman, D. M. *et al.* Medulloblastoma growth inhibition by Hedgehog pathway blockade. *Science* **297**, 1559–1561 (2002).
18. Park, H. L. *et al.* Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. *Development* **127**, 1593–1605 (2000).
19. Weaver, M., Yingling, J. M., Dunn, N. R., Bellusci, S. & Hogan, B. L. Bmp signaling regulates proximal-distal differentiation of endoderm in mouse lung development. *Development* **126**, 4005–4015 (1999).
20. Borges, M. *et al.* An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature* **386**, 852–855 (1997).
21. Kenney, A. M. & Rowitch, D. H. Sonic hedgehog promotes G(1) cyclin expression and sustained cell cycle progression in mammalian neuronal precursors. *Mol. Cell Biol.* **20**, 9055–9067 (2000).
22. Dahmane, N. & Ruiz-i-Altaba, A. Sonic hedgehog regulates the growth and patterning of the cerebellum. *Development* **126**, 3089–3100 (1999).
23. Wechsler-Reya, R. J. & Scott, M. P. Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron* **22**, 103–114 (1999).
24. Johnson, D. H. Management of small cell lung cancer*: current state of the art. *Chest* **116**, 525S–530S (1999).
25. Hogan, B. L., Bedington, R., Constantini, F. & Lacy, E. *Manipulating the Mouse Embryo* (Cold Spring Harbor Press, Plainview, 1994).
26. Chang, D. T. *et al.* Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development* **120**, 3339–3353 (1994).
27. Wang, B., Fallon, J. F. & Beachy, P. A. Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* **100**, 423–434 (2000).
28. Sriuranpong, V. *et al.* Notch signaling induces rapid degradation of achaete-scute homolog 1. *Mol. Cell Biol.* **22**, 3129–3139 (2002).
29. Nakakura, E. K. *et al.* Mammalian Scratch: a neural-specific Snail family transcriptional repressor. *Proc. Natl Acad. Sci. USA* **98**, 4010–4015 (2001).
30. Park, B. H., Vogelstein, B. & Kinzler, K. W. Genetic disruption of PPARdelta decreases the tumorigenicity of human colon cancer cells. *Proc. Natl Acad. Sci. USA* **98**, 2598–2603 (2001).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We acknowledge M. Scott for the donation of the *Ptch-LacZ* mice; A. Joyner for the Flag-Gli1 expression vectors; J. Chen for the gift of KAAD cyclopamine; J. Taipale, S. Karhadkar, B. Nelkin and K. Schuebel for discussions; and E. Gabrielson for help in obtaining lung cancer tissue. We also thank K. Young and L. Meszler and the Sidney Kimmel Comprehensive Cancer Center Cell Imaging Facility for technical assistance. This work is supported by the Flight Attendant Medical Research Institute and the NCI/SPORC. P.A.B. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to D.N.W.
(e-mail: nwatkns@jhmi.edu).

Links between signal transduction, transcription and adhesion in epithelial bud development

Colin Jamora*, Ramanuj DasGupta*†, Pawel Kocieniewski* & Elaine Fuchs*

* Howard Hughes Medical Institute, Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, New York 10021, USA

The morphogenesis of organs as diverse as lungs, teeth and hair follicles is initiated by a downgrowth from a layer of epithelial stem cells^{1,2}. During follicular morphogenesis, stem cells form this bud structure by changing their polarity and cell–cell contacts. Here we show that this process is achieved through simultaneous receipt of two external signals: a Wnt protein to stabilize β -catenin, and a bone morphogenetic protein (BMP) inhibitor to produce Lef1. β -Catenin then binds to, and activates, Lef1 transcription complexes that appear to act uncharacteristically by downregulating the gene encoding E-cadherin, an important component of polarity and intercellular adhesion. When either signal is missing, functional Lef1 complexes are not made, and E-cadherin downregulation and follicle morphogenesis are impaired. In *Drosophila*, E-cadherin can influence the plane of cell division and cytoskeletal dynamics³. Consistent with this notion, we show that forced elevation of E-cadherin levels block invagination and follicle production. Our findings reveal an intricate molecular programme that links two extracellular signalling pathways to the formation of a nuclear transcription factor that acts on target genes to remodel cellular junctions and permit follicle formation.

During skin development, signals from adjacent epithelial and mesenchymal cells instruct select ectodermal cells to form hair follicle buds. In turn, each bud signals to a small group of underlying mesenchymal cells to condense^{1,2}. Once the bud proliferates to form a larger bulb (matrix), it encases this dermal condensate (papilla), and further differentiates into the cells of the hair shaft (Fig. 1a)¹. Recent evidence suggests that Wnt signalling is involved in this process at a time that correlates with bud-specific patterns of upregulation of P-cadherin and downregulation of E-cadherin^{4–9}. Cadherins form the transmembrane core of adherens junctions (AJs) by bridging to α -catenin and the cytoskeleton through β -catenin, a protein which on its own is prone to degradation^{10,11}. β -Catenin's degradation machinery is transiently suppressed by Wnt signalling. This renders β -catenin a new-found stability and function, binding to and activating members of the

† Present address: Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA.

Lef1/Tcf family of DNA binding proteins^{11–13}. Whether there is underlying functional significance to β -catenin's link between adhesion and transcription is an issue that we now address.

Wnts are expressed in ectodermal buds^{14,15}, and are prime candidates to stabilize β -catenin at these sites. We confirmed this by testing the ability of the canonical skin Wnt3a to generate nuclear β -catenin in mouse keratinocytes. Keratinocytes exposed to Wnt3a-conditioned media displayed an ~ 7 times increase in β -catenin, as judged by immunoblot and densitometry analysis (Fig. 1b). This increase was paralleled by accumulation of β -catenin in $\sim 85 \pm 5\%$ of the nuclei of treated cells (Fig. 1c).

Wnt-treated cultures did not express appreciable Lef1, suggesting the need for additional signalling molecules to induce a DNA binding protein for β -catenin activation. As bud formation *in vivo* requires a mesenchymal cue¹, we searched for candidates expressed by developing dermal condensates. Epithelial cells and mesenchymal cells within follicle buds express BMP2 and BMP4 (refs 16, 17), while only mesenchymal cells express their inhibitor, noggin¹⁸. Polymerase chain reaction with reverse transcription (RT-PCR) and western analyses revealed that *in vitro*, keratinocytes express BMPs, but not noggin (Fig. 1d). When exposed to noggin-

conditioned media, Lef1 was induced, and localized to the nucleus (Fig. 1b, c).

Together, Wnt and noggin promoted transactivation of the *TOPFLASH* reporter gene, which regulated luciferase expression through Lef/Tcf binding sites in the *TOP* promoter¹² (Fig. 1e). When the Tcf/Lef1 binding sites were mutated (*FOPFLASH*), promoter activity was abolished, as it was when noggin was absent and BMPs were present. Additionally, although some stabilization of cytoplasmic β -catenin occurred naturally in keratinocyte cultures, transactivation was clearly enhanced by Wnt3a. Thus, *in vitro*, noggin and Wnt3a appeared to act in concert to generate transcriptionally competent Lef1 complexes.

Noggin and Wnts also appeared to be essential for this process *in vivo* (Fig. 1f). In skin from embryonic day (E) 16.5, nuclear Lef1 and β -catenin could be seen in most follicles except for the mature guard follicles developing at E13.5 independently of Lef1. Lef1-positive hair buds also expressed β -galactosidase under the control of the *TOP* promoter⁷. In contrast, *noggin*-null (*Nog*^{-/-}) mice lacked two-thirds of E16.5 buds¹⁸, and those that formed exhibited cytoplasmic β -catenin and little or no Lef1 or TOPGAL activity (Fig. 1f). When *Nog*^{-/-}/*TOPGAL* mice were mated on a back-

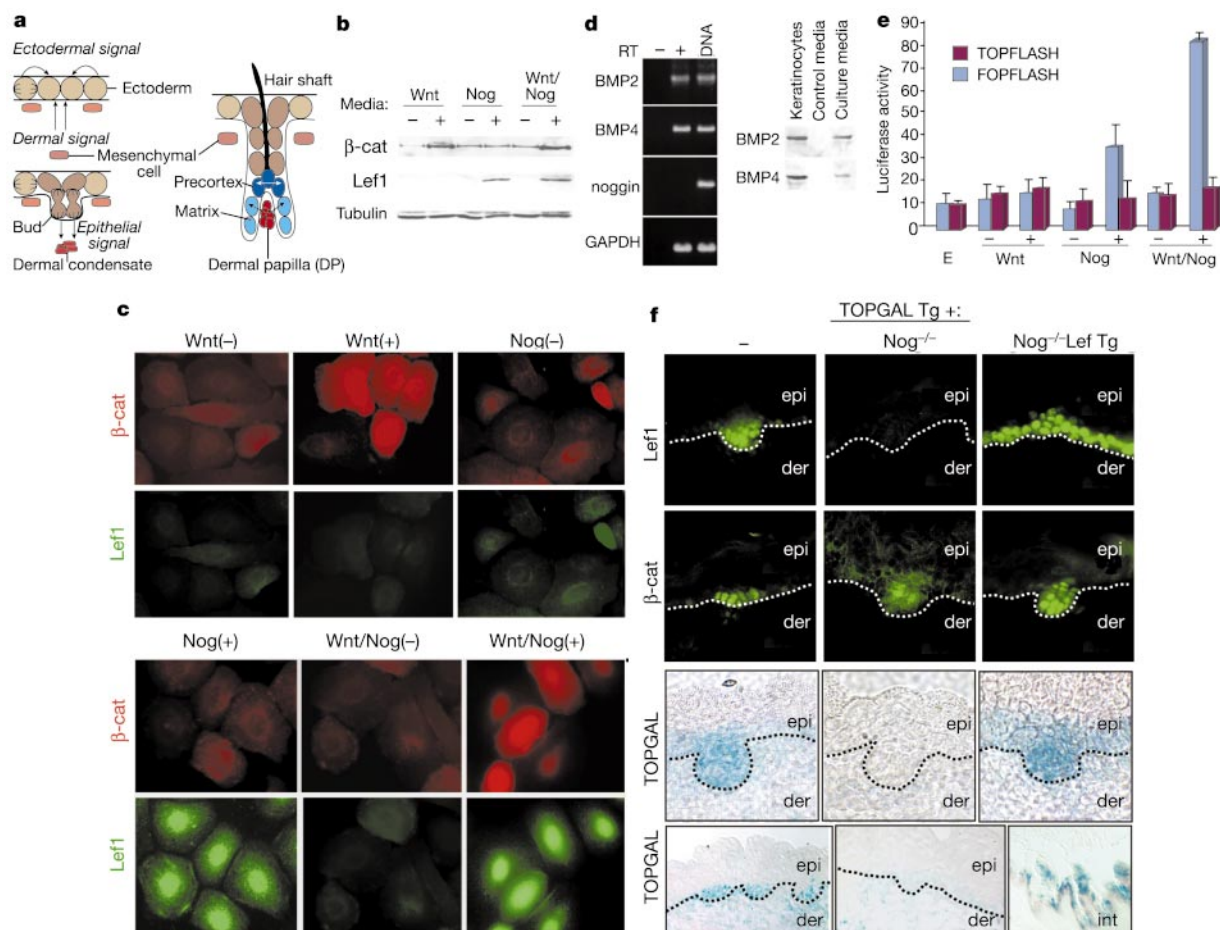


Figure 1 Wnt and noggin induce a transcriptionally competent Lef1 complex that is absent in *Nog*^{-/-} mice. **a**, Cell-cell signalling (denoted by arrows) in developing follicles: noggin is expressed by mesenchyme and Wnts by ectoderm. **b**, SDS-PAGE immunoblots showing Wnt3a-induced β -catenin upregulation and noggin-induced Lef1 induction in lysates from keratinocytes treated with control (-) or conditioned (+) media. **c**, Immunofluorescence of keratinocytes, revealing nuclear Wnt3a-induced β -catenin and noggin-induced Lef1. **d**, RT-PCR and immunoblot analyses revealing BMP2 and BMP4, but not noggin, in keratinocytes. **e**, Effect of Wnt3a and/or noggin on *TOPFLASH* (test) and *FOPFLASH* (control) in transiently transfected keratinocytes. E, low calcium media.

f, Dependency of nuclear β -catenin, Lef1 and *TOPGAL* expression on *noggin* *in vivo*. E16.5 skins from the knockout/transgenic (Tg) mice as indicated at top. Note that in epithelial buds lacking noggin, neither Lef1 nor *TOPGAL* are expressed and β -catenin is diffuse; these features are rescued on the K14-Lef1 transgenic background. Dotted lines denote mesenchymal-epithelial boundaries between epidermis (epi) and dermis (der). Bottom row are low-magnification views; lower right is of *Nog*^{-/-} intestinal epithelium, which uses Tcf4, rather than Lef1, to regulate Wnt signalling⁷. For **b**, **c**, **f**, antibodies or *TOPGAL* activity assays are denoted at left.

ground of *K14-Lef1*⁵ to force Lef1 expression, nuclear β -catenin and *TOPGAL* expression were restored in ~80% of the E16.5 buds (Fig. 1f). This said, follicles were not restored to wild-type levels, indicating that some of the altered genes and morphology associated with *Nog*^{-/-} follicles are not under control of Lef1/ β -catenin.

When cells of embryonic epidermis (epi) reorient to form an epithelial bud, they switch from E-cadherin to P-cadherin based AJs (Fig. 2a, b). This timing coincided with Lef1 and *TOPGAL* activation (Fig. 2c, d). At the mesenchymal–epithelial interface, the cadherin switch persisted throughout follicle development, and was still in Lef1-positive, stem cell progeny of adult follicles (Fig. 2e–g). As in buds, adult hair precursor cells were not only E-cadherin dim and P-cadherin bright, but also active for nuclear Lef1, β -catenin and *TOPGAL*^{7,19}. These results were interesting in light of findings by Shimamura *et al.*²⁰, who noted that in developing mouse brain, E-cadherin was absent wherever Wnt-1 was expressed. A correlation between reduced E-cadherin and elevated nuclear β -catenin has also been observed *in vitro* and in human cancers²¹.

In situ hybridization revealed that the cadherin switch was regulated at the messenger RNA level (Fig. 2h, i). Both *cadherin* promoters harbour multiple sequence motifs corresponding to the optimal Lef1/Tcf binding site, and the *E-cadherin* promoter has been shown to bind Lef1²². Although P-cadherin appeared unaffected by a *Nog*-null or *Lef1*-null background, *E-cadherin* mRNA and protein failed to be downregulated (Fig. 2j–m), a feature which was restored when mice were bred on the *K14-Lef1* background

(Fig. 2n). These data reveal the *E-cadherin* gene as a candidate for negative repression by noggin-induced Lef1.

To further assess the extent to which *E-cadherin* downregulation correlated specifically with follicle downgrowth, we analysed E16.5 skin null for *sonic hedgehog* (*Shh*), which when absent impairs follicle development soon after Lef1/ β -catenin activation¹⁴. Arrested *Shh*^{-/-} hair buds were positive for P-cadherin, and suppressed E-cadherin at both protein (shown) and mRNA (not shown) levels (Fig. 2o, p). Taken together, these data place *noggin* involvement earlier than *Shh*, and specifically required for *E-cadherin* downregulation.

A priori, the negative effects of Lef1 on *E-cadherin* mRNA expression could be independent of β -catenin, a view consistent with how Tcf/Lef family members are known to repress target genes¹². To test this possibility, we examined E-cadherin expression in skin from a mouse expressing a constitutively stable β -catenin. In interfollicular epidermis, the transgene (*K14 Δ N β -catenin*) elicits *de novo* follicle-like downgrowths⁶, which display evidence of Wnt-responsive gene transactivation on a *TOPGAL* background⁷. E-cadherin downregulation was consistently observed at Lef1-positive sites of *K14 Δ N β -catenin* induced epithelial invaginations (Fig. 2q, r). A complementary study on *Wnt-1*-null mice reported a broadening of the E-cadherin pattern in developing brain tissue²⁰. Our findings now uncover a dual and atypical importance of both stabilized β -catenin and Lef1 in repressing *E-cadherin* expression.

To begin to address whether the changes we observed in *E-cadherin* mRNA expression are reflected at the transcriptional level,

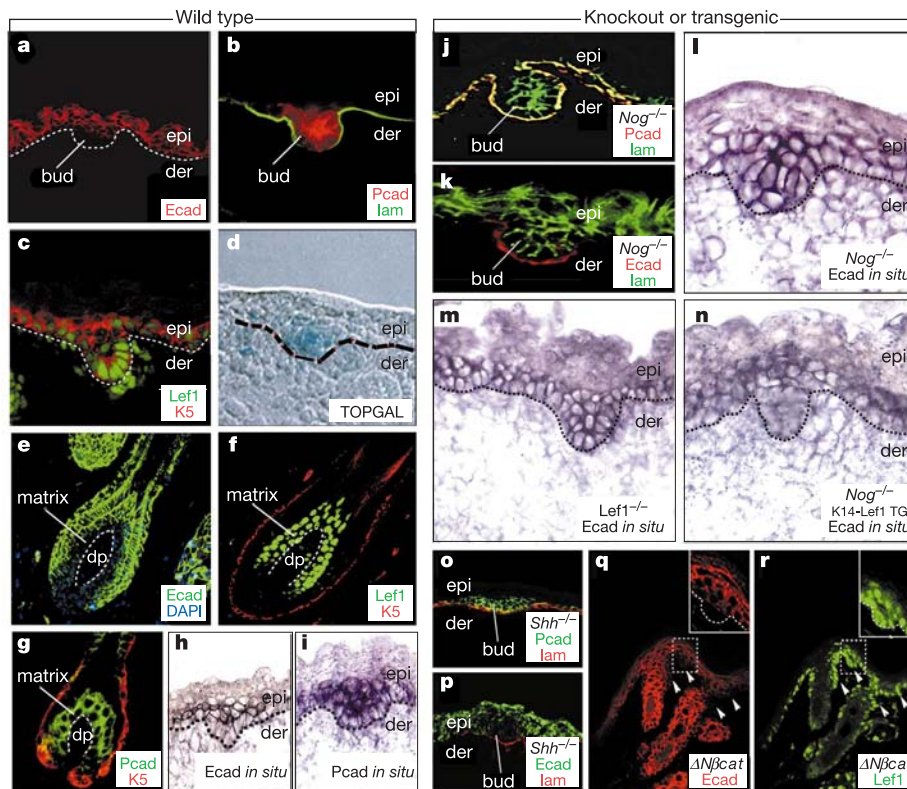


Figure 2 During follicle morphogenesis and cycling, skin stem cells change cadherin expression in a fashion dependent upon noggin and Wnt signalling. Wild-type skins from *TOPGAL* reporter mice at E16.5 (**a–d, h, i**) or adult (**e–g**) (left) were processed for double immunofluorescence, *TOPGAL* (β -galactosidase) activity or *in situ* hybridizations, using the markers indicated (colour coding denotes secondary antibodies). Wherever noggin and Wnt signalling were active, *E-cadherin* was downregulated, and nuclear Lef1 was present, indicating β -catenin activated Lef1 complexes. This was true even for *Shh*-null

skin (**o, p**), arrested at the bud stage, and for the Lef1-positive epithelial invaginations (arrowheads) of *K14- Δ N β cat* transgenic (Tg) skin, expressing stable β -catenin (**q, r**). Without *noggin*, few buds formed and *E-cadherin* remained high (**j–m**), but this was rescued by *K14-Lef1* (**n**). lam, laminin 5 antibody, demarcating the epithelial–mesenchymal boundary (elsewhere delineated by dotted lines); K5 antibody, marking the basal layer of epidermis and outer root sheath of the hair follicle. epi, embryonic epidermis; der, dermis; DP, dermal papilla.

we first cloned 6.5 kilobases (kb) of 5' murine *E-cadherin* gene sequence, and engineered a mutation in the promoter that had previously been shown by gel shift assays to bind recombinant Lef1²². When tested in primary mouse keratinocytes, both wild-type and mutant promoters yielded comparable β -galactosidase reporter gene expression, indicating that the Lef1/Tcf site is not required for promoter activity (Fig. 3a).

Transfection of either stable β -catenin or Lef1 on their own did not appreciably affect the activity of the *E-cadherin* promoter. In contrast, Lef1 and stabilized β -catenin in combination markedly suppressed *E-cadherin* promoter activity in a manner dependent upon the Lef1/Tcf binding site (Fig. 3a). Similar repression was observed when noggin and Wnt 3a together were added to the keratinocyte cultures (Fig. 3b).

To test whether Wnt-activated Lef1 complexes bind to the endogenous *E-cadherin* promoter, we conducted chromatin immunoprecipitation analyses (ChIP) on Lef1-expressing keratinocytes. Only when these cells were exposed to Wnt3A were we able to specifically precipitate 120-base-pair (bp) chromatin fragments of the *E-cadherin* promoter containing the Lef1 binding site (Fig. 3c). In contrast to the repressive effects on the *E-cadherin* promoter, Wnt and noggin together activated the murine *HK1-hair keratin* promoter, previously identified as a bona fide Lef1/ β -catenin responsive target in keratinocytes (Fig. 3d)¹⁹.

Our findings suggest that Lef1's ability to function as repressor and activator in the same Wnt-treated cells depends upon the context of responsive promoter elements. Located 3' from the Lef1 motif in the *E-cadherin* promoter is an E-box sequence, which in various cell lines serves as the binding and regulatory site for the Snail family of transcriptional repressors^{23,24}. Mutation

of the Lef1 binding site did not interfere with Snail's ability to repress the *E-cadherin* promoter (Fig. 3a). Moreover, these two sites together functioned additively to repress wild-type *E-cadherin* promoter activity by up to 70% of its normal activity in keratinocytes (Fig. 3a).

Despite the presence of functional *in vitro* binding sites for Snail and Lef1, the *E-cadherin* promoter might still be regulated *in vivo* by indirect pathways involving these factors. Irrespective of mechanism, the independent repressor action of Snail was intriguing given that the large guard hair follicles that develop on both the *noggin*-null and *Lef1*-null genetic backgrounds still exhibited *E-cadherin* downregulation (not shown). These data reveal *E-cadherin* downregulation as a common thread among the waves of follicle morphogenesis and expose the existence of multiple mechanisms to govern this downregulation.

To test the functional importance of *E-cadherin* downregulation in hair follicle morphogenesis, we engineered transgenic mice expressing elevated levels of an epitope-tagged *E-cadherin*. Previous studies have shown that the addition of a carboxy-terminal tag does

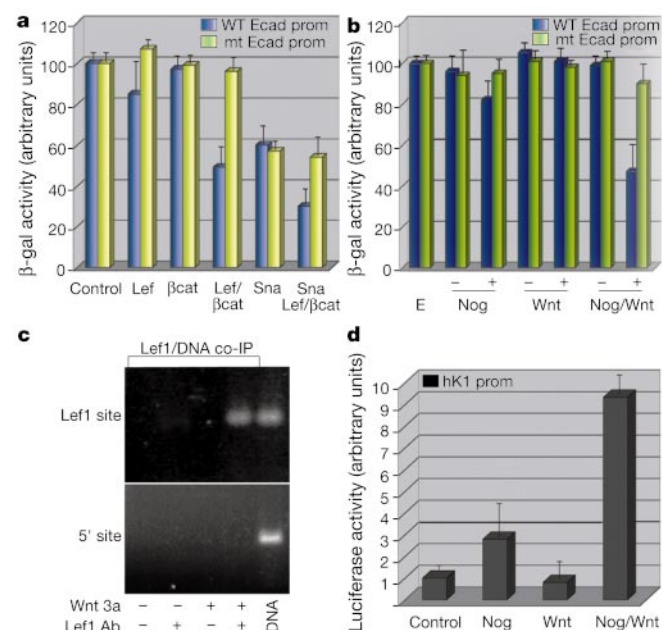


Figure 3 Lef1 and β -catenin transcriptionally downregulate *E-cadherin* *in vitro*. **a, b, d**, Promoter activity assays on extracts from keratinocytes treated without (**a**) or with (**b, d**) control (E) or conditioned media as indicated. Cells were co-transfected with *E-cadherin*- β -galactosidase (WT (wild-type) or mt (mutant) Lef1-site; **a, b**) or *HK1-luciferase* (**d**)¹⁹ $\pm$ K14-expression vectors encoding Lef1, β cat (Δ N β cat), Snail (Snail), or control (empty vector). **c**, Chromatin IP analyses of Lef1-expressing keratinocytes $\pm$ Wnt-conditioned medium. Fragmented, crosslinked DNA from chromatin immunoprecipitated with Lef1 or control antibody was subjected to PCR with primers encompassing the known Lef1 site or control sequence in the *E-cadherin* promoter. DNA, genomic control.

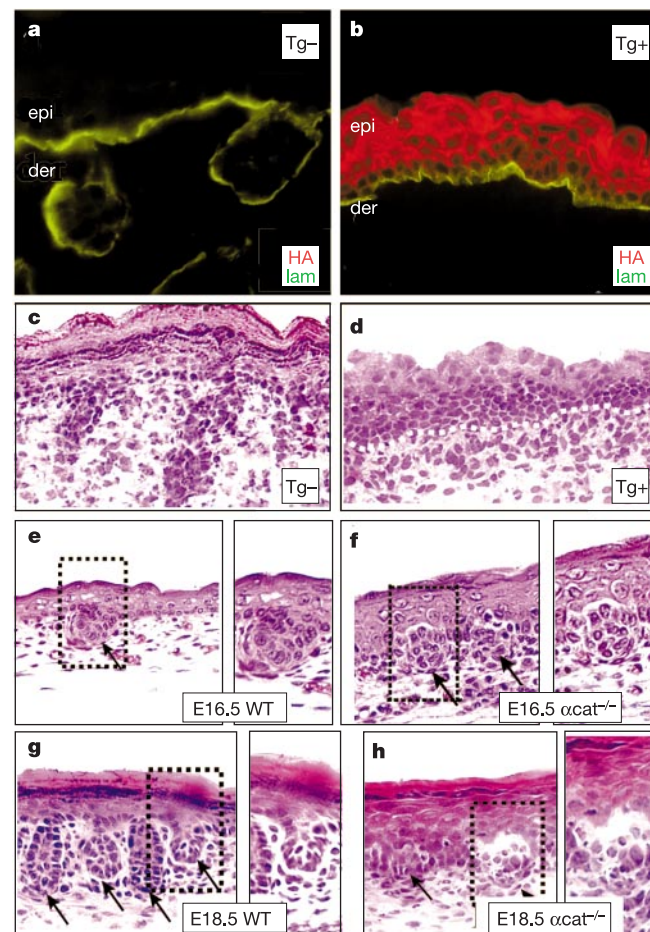


Figure 4 Levels of adherens junction proteins influence follicle morphogenesis. **a-d**, Immunofluorescence and haematoxylin-eosin staining of frozen back-skin sections from a representative newborn mouse mosaic for the *K14-Ecadherin-HA* transgene. Transgene-negative (Tg -) and positive (Tg +) regions were identified by anti-HA (HA), and anti-laminin 5 (lam) delineated the basement membrane. **e-h**, Haematoxylin-eosin staining of back-skin sections from wild-type and α -catenin K14-Cre conditional-null (α cat^{-/-}) embryos. Boxed areas are shown at higher magnification to right of each frame. Note that in α -catenin-null skin, follicle buds (arrows) begin to form at E16.5, but cellular disorganization arrests their downgrowth by E18.5.

not interfere with E-cadherin's ability to form intercellular junctions²⁵. Several of the newborn animals harbouring the *K14-Ecadherin-HA* transgene were sickly, and were killed shortly after birth. Immunofluorescence analyses of frozen newborn skin sections revealed a mosaic pattern of anti-HA negative (Fig. 4a) and positive (Fig. 4b) domains. Signs of the characteristic E-cadherin down-regulation were seen in the non-transgenic (Tg⁻) epithelium, but were not detected in transgene positive (Tg⁺) areas. Haematoxylin and eosin staining confirmed that the most notable morphological difference between Tg⁻ and Tg⁺ regions was the paucity of hair follicles in Tg⁺ skin (Fig. 4c, d).

Skin conditionally null for *a-catenin* (by *K14-Cre*) also lacked proper AJs²⁶. At E16.5, developing placodes were visible (Fig. 4e, f), but they were severely arrested by birth (Fig. 4g, h). Correspondingly, sebaceous glands failed to develop. These findings demonstrate the deleterious consequences of too few as well as too many AJ proteins in hair follicle morphogenesis, and provide compelling evidence that dynamic changes in *AJ* gene expression may be a key step in morphogenesis. In this regard, it seems particularly relevant that either complete loss or overexpression of E-cadherin can also impair formation of intestinal epithelia and mammary glands^{27,28}. In contrast, a reduction in AJs is characteristic of epithelial cancers, whose cellular masses bear some resemblance to aberrant bud formation²¹. The extent to which these effects go beyond simple adhesion is at present unknown.

Our findings shed new light on the molecular mechanism underlying the early steps of epithelial bud development, a process that is anticipated to have implications that extend beyond the hair follicle. Inverse correlations between E-cadherin and Wnt signalling have been noted in various tissues and organs, during development and in cancers. In some human epithelial cancers, *E-cadherin* gene mutations seem to promote not only malignant invaginations but also Wnt signalling, as judged by the detection of nuclear β -catenin in the tumour tissue²¹. Our studies now provide evidence for a converse mechanism, namely nuclear β -catenin/Lef1 mediated suppression of E-cadherin expression during normal follicle development. The downregulation of E-cadherin may in turn perpetuate the Wnt signalling pathway by increasing the pool of transcriptionally competent β -catenin¹⁰.

Although Lef1/Tcf factors are known to function as repressors, β -catenin involvement is typically associated with converting these proteins from repressors to activators¹². In this regard, our *in vitro* results on the *E-cadherin* promoter are surprising, as they suggest a role for β -catenin in mediating the repressive effects of Lef1 on the *E-cadherin* promoter. One possibility is that the atypical context of the Lef1 binding site in the *E-cadherin* promoter may recruit other cofactors, such as Snail, to endow β -catenin stimulated Lef1 with this unusual repressor activity. Although we have shown that the Snail and Lef1 binding sites can function independently, Lef1's ability to bend DNA still affords a means of bringing these complexes into contact. Additionally, β -catenin's ability to seemingly act through both chromatin-dependent¹² and independent¹³ mechanisms to modulate Lef1's action hints at unexplored avenues for how these negative activities might be coordinated *in vivo*. But our results do not preclude the possibility that *in vivo*, the ability of Lef1 and β -catenin to downregulate *E-cadherin* mRNA expression may be indirect, perhaps through transcriptionally activating one of the E-box repressors, such as Snail, that bind to this promoter²⁹. Once transcriptionally downregulated, E-cadherin's relatively short half-life and multiple modes of posttranslational regulation¹⁰ may facilitate the ensuing changes that accompany follicle morphogenesis.

We have found that in noggin-regulated waves of Lef1-expressing follicles, Wnts appear to collaborate by stabilizing β -catenin so that it can bind and activate a transcription factor that participates in finely balancing the levels of E-cadherin to sculpt the epithelial bud. The sources of noggin and Wnt offer an additional level of

regulation in the process. Whereas canonical Wnts have been found in skin epithelium^{14,15}, noggin is mesenchymal¹⁸, and BMPs and their surface receptors are expressed by both mesenchyme and epithelium^{16,17}. Our findings suggest a model in which embryonic skin epithelial stem cells require simultaneous inputs of stimulatory and inhibitory signals from multiple neighbouring cell types for the purpose of producing an activated transcription factor able to remodel AJ gene expression and form a follicle bud (see Supplementary Information). Such a mechanism furnishes an exquisite level of governance by converging signals to specify stem cell lineage. □

Methods

Plasmid construction

TOPFLASH and *FOPFLASH* were gifts (H. Clevers). The *E-cadherin* promoter was generated by PCR with primers (Celera Genomics database) and a BAC clone as template. Mutation of the Lef1 binding site (-242 to -233) was achieved using the QuickChange Site-Directed Mutagenesis Kit (Stratagene). Promoter fragments were subcloned into pGL3-basic (Promega; luciferase) or pNASS β (M) (β -galactosidase)⁷. *HA-tagged Snail* cDNA (A. Garcia de Herreros) was subcloned into the *Bam*HI/*Not*I sites of the *K14-cassette*⁶. *E-cadherin-HA* cDNA was generated by PCR amplification of pKS + UM plasmid (R. Kemler) with a forward primer at the *Xho*I site (3' end), and a reverse primer containing the HA tag, a stop codon and *Xba*I site. E-cadherin-HA localized to intercellular junctions in keratinocytes.

Nuclear Lef1 and β -catenin generated by noggin and Wnt3a conditioned media

Keratinocytes from newborn mouse skin were cultured in low calcium E-media⁷, and then either treated for (1) 12 h with control-media (Nog(-)) or conditioned-media (Nog(+)) from a noggin-secreting cell line (R. Harland), or (2) 5 h with control-media (Wnt(-)) or conditioned-media (Wnt(+)) from a Wnt3A-secreting cell line (S. Takada). In some cases, 7 h noggin/control-media was replaced with Nog/Wnt conditioned or Nog(-)/Wnt(-) media for 5 h. In all cases, EGTA was added to 5 mM and control/conditioned media was at a final dilution of 1:5.

For harvesting, cells were washed twice with PBS and lysed in RIPA buffer (1% Triton X100, PBS, 10 mM EDTA, 150 mM NaCl, 1% sodium deoxycholate, 0.1% SDS, protease inhibitors). After determining protein concentrations (BCA, Pierce) for equal loadings, SDS-PAGE and immunoblot analyses were performed with mouse anti-tubulin (Sigma), rabbit anti-Lef1 (UC72), and mouse anti- β -catenin (Sigma). HRP-conjugated secondary-antibodies were followed by enhanced chemiluminescence (Amersham).

BMP/Noggin expression analyses

RNAs were purified by Trizol-extraction of keratinocytes (Invitrogen) and spin-column chromatography (Qiagen). Reverse transcription was performed with a Superscript kit (Invitrogen). *BMP2*, *BMP4*, *noggin*, and *GAPDH* mRNAs were detected by RT-PCR. For BMP immunodetection, 1 ml of spent-keratinocyte or control media were precipitated with two volumes of acetone, while keratinocytes were pelleted directly by centrifugation. Samples were subjected to 10% SDS-PAGE and immunoblotting with anti-BMPs (R&D Systems).

Transfections

FUGENE6 (Roche) was used to transfect newborn skin keratinocytes with expression vectors. For β -galactosidase reporters, *CMV-luciferase* was used to control for transfection efficiency, and for luciferase reporters, *CMV- β -galactosidase* was used. After 24 h, cells were treated with Wnt and/or noggin media as described. β -Galactosidase activity was measured with the Galacto-Lite-Assay Kit (Tropix Inc.) and luciferase activity by the Dual-Luciferase Kit (Promega). For standardizations, the activity in the transfected lysate of control cells was assigned an arbitrary value.

Animals

noggin and *Shh* null mice (A. McMahon) and *Lef1* null mice (R. Grosschedl) were gifts. α -Catenin conditional null, *TOPGAL*, *K14-Lef1* and *K14- Δ N β cat* animals have been described^{5-7,27}. *K14-EcadherinHA* transgenic mice were generated in the laboratory.

In situ hybridization

Digoxigenin-labelled probes were synthesized with the DIG-RNA labelling kit and detected with Anti-Digoxigenin AP (both from Roche). Antisense *E-cadherin* cRNA was generated from T7 RNA polymerase transcription of a *Bgl*II-*Xho*I extracellular domain fragment that had been subcloned into pCRII (Invitrogen) and linearized with *Hind*III. Sense probe was generated by *Xho*I digestion and Sp6 RNA polymerase transcription. 3' UTR *P-cadherin* cRNAs were generated by subcloning a 379 bp PCR product into pCRII. Probes were applied to 15- μ m sections of frozen, optimal cutting temperature (OCT) compound-embedded tissue and processed as described⁷.

Immunohistochemistry and tissue analysis

Sections (10 μ m) of frozen embryos or derived-keratinocytes were fixed in 4% paraformaldehyde and processed for indirect immunofluorescence. Primary antibodies used were: rabbit anti-Lef1 (UC54), rat anti-E-cadherin (M. Takeichi), rat anti-P-cadherin (Zymed), guinea pig anti-K5, rat anti-HA (Roche) and rabbit anti-laminin5 (8LN5; R. Burgeson). Secondary FITC (fluorescein isothiocyanate) or Texas Red conjugated antibodies (Jackson Labs) were diluted 1:100. For detection of nuclear β -catenin, whole

embryos were fixed overnight in 4% PFA, dehydrated, embedded in paraffin, and rehydrated before probing with mouse anti- β -catenin antibody (Clone 15B8, Sigma). Expression of TOPGAL was determined by X-gal staining of frozen embryo sections (7.5 μ m) embedded in OCT and fixed in 0.1% glutaraldehyde. Histological analysis of embryonic skin was performed by staining with haematoxylin and eosin (Richard Allan Scientific).

ChIP and PCR

Protein-DNA complexes were crosslinked in whole cells with 1% formaldehyde, followed by sonication to fragment genomic DNA to a mid-range of 600 bp (protocols from H. Singh). DNA from anti-Lef1 immunoprecipitation was subjected to PCR using Taq-polymerase (Promega) and primers specific for a 123-bp sequence encompassing the Lef1 binding site (Lef1 site; 5' CAAAGAAAATAAAACATAAGAAAC3'; 5' TCCTATTCCACGGTCGTTGCG3')²² and a site ~2 kb upstream (5' site; 5' AGCACTCTATAGATGAGGC3', 5' TACTAAGGC-CAAAACAATCACTG3'). PCR was performed with 40 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 45 s, 72 °C for 30 s. PCR products were separated on 1.5% agarose gels and visualized with ethidium bromide.

Received 6 November 2002; accepted 15 January 2003; doi:10.1038/nature01458.

- Hardy, M. H. The secret life of the hair follicle. *Trends Genet.* **8**, 55–61 (1992).
- Hogan, B. L. Morphogenesis. *Cell* **96**, 225–233 (1999).
- Jan, Y.-N. & Jan, L.-Y. Asymmetric cell division in the Drosophila nervous system. *Nature Rev. Neurosci.* **2**, 772–779 (2001).
- van Genderen, C. et al. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in Lef-1-deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
- Zhou, P., Byrne, C., Jacobs, J. & Fuchs, E. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes Dev.* **9**, 700–713 (1995).
- Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. De novo hair follicle morphogenesis and hair tumors in mice expressing a truncated β -catenin in skin. *Cell* **95**, 605–614 (1998).
- DasGupta, R. & Fuchs, E. Multiple roles for activated Lef/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
- Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. β -catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105**, 533–545 (2001).
- Hirai, Y., Nose, A., Kobayashi, S. & Takeichi, M. Expression and role of E- and P-cadherin adhesion molecules in embryonic histogenesis. II. Skin morphogenesis. *Development* **105**, 271–277 (1989).
- Gumbiner, B. M. Regulation of cadherin adhesive activity. *J. Cell. Biol.* **148**, 399–404 (2000).
- Daniels, D. L., Eklof-Spink, K. & Weis, W. I. β -catenin: molecular plasticity and drug design. *Trends Biochem. Sci.* **26**, 672–678 (2001).
- Brantjes, H., Barker, N., van Es, J. & Clevers, H. TCF: Lady Justice casting the final verdict on the outcome of Wnt signalling. *Biol. Chem.* **383**, 255–261 (2002).
- Chan, S.-K. & Struhl, G. Evidence that Armadillo transduces Wingless by mediating nuclear export or cytosolic activation of pangolin. *Cell* **111**, 265–280 (2002).
- St-Jacques, B. et al. Sonic hedgehog signaling is essential for hair development. *Curr. Biol.* **8**, 1058–1068 (1998).
- Reddy, S. et al. Characterization of Wnt gene expression in developing and postnatal hair follicles and identification of Wnt5a as a target of Sonic hedgehog in hair follicle morphogenesis. *Mech. Dev.* **107**, 69–82 (2001).
- Kratohwil, K., Dull, M., Farinas, I., Galceran, J. & Grosschedl, R. Lef1 expression is activated by BMP-4 and regulates inductive tissue interactions in tooth and hair development. *Genes Dev.* **10**, 1382–1394 (1996).
- Kulesa, H., Turk, G. & Hogan, B. L. Inhibition of BMP signaling affects growth and differentiation in the anagen hair follicle. *EMBO J.* **19**, 6664–6674 (2000).
- Botchkarev, V. A. et al. Noggin is a mesenchymally derived stimulator of hair-follicle induction. *Nature Cell Biol.* **1**, 158–164 (1999).
- Merrill, B. J., Gat, U., DasGupta, R. & Fuchs, E. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev.* **15**, 1688–1705 (2001).
- Shimamura, K., Hirano, S., McMahon, A. P. & Takeichi, M. Wnt-1-dependent regulation of local E-cadherin and α -N-catenin expression in the embryonic mouse brain. *Development* **120**, 2225–2234 (1994).
- Conacci-Sorrell, M., Zhurinsky, J. & Ben-Ze'ev, A. The cadherin-catenin adhesion system in signaling and cancer. *J. Clin. Invest.* **109**, 987–991 (2002).
- Huber, O. et al. Nuclear localization of β -catenin by interaction with transcription factor Lef-1. *Mech. Dev.* **59**, 3–10 (1996).
- Cano, A. et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biol.* **2**, 76–83 (2000).
- Battle, E. et al. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature Cell Biol.* **2**, 84–89 (2000).
- Adams, C. L., Chen, Y. T., Smith, S. J. & Nelson, W. J. Mechanisms of epithelial cell-cell adhesion and cell compaction revealed by high-resolution tracking of E-cadherin-green fluorescent protein. *J. Cell Biol.* **142**, 1105–1119 (1998).
- Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of α -catenin in skin. *Cell* **104**, 605–617 (2001).
- Hermiston, M. L., Wong, M. H. & Gordon, J. I. Forced expression of E-cadherin in the mouse intestinal epithelium slows cell migration and provides evidence for nonautonomous regulation of cell fate in a self-renewing system. *Genes Dev.* **10**, 985–996 (1996).
- Boussadia, O., Kutsch, S., Hierholzer, A., Delmas, V. & Kemler, R. E-cadherin is a survival factor for the lactating mouse mammary gland. *Mech. Dev.* **115**, 53–62 (2002).
- Kolligs, F. T. et al. ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with β -catenin defects and promotes neoplastic transformation. *Cancer Cell* **1**, 145–155 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank L. Degenstein and J. Fan for transgenic technology; B. Merrill and M. Perez-Moreno for comments on the manuscript; W. Lowry and H. Rhee for sharing experimental expertise; M. Takeichi and R. Burgeson for antibodies; A. McMahon and R. Grosschedl for knockout mice; R. Kemler, A. Garcia de Herreros and H. Clevers for cDNAs and vectors; R. Harland and S. Takada for cell lines. C.J. is supported by the Helen Hay Whitney foundation. R.D. was a graduate student at the University of Chicago, where a portion of this work was conducted. This work was supported by the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.F. (e-mail: fuchs@rockefeller.edu).

Segregation of receptor and ligand regulates activation of epithelial growth factor receptor

Paola D. Vermeer*, Lisa A. Einwalter*, Thomas O. Moninger†, Tatiana Rokhlina*, Jeffrey A. Kern‡, Joseph Zabner* & Michael J. Welsh*§

* Department of Internal Medicine, † Central Microscopy Research Facility, § Department of Physiology and Biophysics, Howard Hughes Medical Institute, Roy J. and Lucille A. Carver College of Medicine, Iowa City, Iowa 52242, USA ‡ University Hospitals of Cleveland, 11100 Euclid Avenue, 610 Wearn, LKSD 5067, Cleveland, Ohio 44106, USA

Interactions between ligands and receptors are central to communication between cells and tissues. Human airway epithelia constitutively produce both a ligand, the growth factor heregulin, and its receptors—erbB2, erbB3 and erbB4 (refs 1–3). Although heregulin binding initiates cellular proliferation and differentiation^{4–7}, airway epithelia have a low rate of cell division⁸. This raises the question of how ligand–receptor interactions are controlled in epithelia. Here we show that in differentiated human airway epithelia, heregulin- α is present exclusively in the apical membrane and the overlying airway surface liquid, physically separated from erbB2–4, which segregate to the basolateral membrane. This physical arrangement creates a ligand–receptor pair poised for activation whenever epithelial integrity is disrupted. Indeed, immediately following a mechanical injury, heregulin- α activates erbB2 in cells at the edge of the wound, and this process hastens restoration of epithelial integrity. Likewise, when epithelial cells are not separated into apical and basolateral membranes ('polarized'), or when tight junctions between adjacent cells are opened, heregulin- α activates its receptor. This mechanism of ligand–receptor segregation on either side of epithelial tight junctions may be vital for rapid restoration of integrity following injury, and hence critical for survival. This model also suggests a mechanism for abnormal receptor activation in diseases with increased epithelial permeability.

The erbB receptors belong to the type I family of tyrosine kinase receptors whose members include erbB1–4 (also called EGFR/HER1 and HER2–4, respectively)^{4,5}. Many gene products, including heregulin, epidermal growth factor (EGF), epiregulin, betacellulin, amphiregulin, TGF- α and epigen serve as ligands for this family of receptors^{4–7}. For example, heregulin binds erbB3 and erbB4. Ligand binding stimulates receptor dimerization, often with erbB2 as one member of the heterodimer. The resulting phosphorylation of erbB cytoplasmic domains activates signal transduction cascades, leading to cellular proliferation and differentiation^{4–7}. Because air-

way epithelia simultaneously express both a ligand, heregulin, and its receptor without manifesting the consequences of receptor activation, we asked how the epithelium controls ligand–receptor interactions.

Reverse transcriptase polymerase chain reaction (RT–PCR; Fig. 1a), microarray analysis (not shown), and western blot analysis (Fig. 1b) showed that well differentiated human airway epithelia expressed heregulin- α but neither heregulin- β nor EGF. They also expressed the receptors erbB1–4. We localized the proteins by immunocytochemistry; Fig. 2a shows an x – z confocal section of the pseudostratified epithelium. Heregulin- α (red) was localized at the apical membrane; some intracellular protein was also observed. The immunostained protein probably represents both membrane-bound pro-heregulin- α and intracellular processed forms. In contrast, erbB2 (green) was present only in the basolateral membrane (Fig. 2a); surface biotinylation confirmed the exclusive basolateral localization of this receptor (Fig. 2i). The physical segregation of heregulin- α and erbB2 was readily appreciated from x – y sections (Fig. 2b, c) taken at the positions indicated in Fig. 2a. We observed a similar staining pattern for erbB1, erbB3 and erbB4 (Fig. 2d–f), but consistent with the western blot analysis, there was no heregulin- β or EGF staining (not shown).

A thin film of airway surface liquid (ASL) covers the apical membrane of differentiated epithelia cultured at the air–liquid interface. We found heregulin- α in ASL, but not in basolateral medium (Fig. 2g). These results are consistent with the immunocytochemical localization, and indicate that heregulin- α , which is synthesized as a transmembrane protein and cleaved to generate a soluble ligand, is released selectively from the apical surface into the ASL. We also found heregulin- α in bronchoalveolar lavage liquid (Fig. 2g), confirming the presence of ligand in ASL *in vivo*. Thus, heregulin- α and its receptors are segregated to opposite sides of the epithelium. These results are consistent with the finding that human proEGF expressed in MDCK cells was present predominantly at the

apical membrane, whereas erbB1 was localized on the basolateral surface⁹.

On the basis of these findings, we reasoned that adding heregulin- α directly to the basolateral surface would activate erbB2. Because the first step following ligand binding is receptor phosphorylation^{4,6}, we examined receptor activation by using an antibody specific to phosphorylated erbB2. As predicted, adding heregulin- α to the basolateral surface phosphorylated erbB2 (Fig. 2h). In contrast, addition of apical heregulin- α had no effect on receptor phosphorylation. To determine whether tight junctions segregated the ligand from its receptor, we opened the junctions by Ca^{2+} chelation; once the junctions were permeabilized, apically added heregulin- α phosphorylated erbB2 (Fig. 2h). In the absence of apically added heregulin- α , Ca^{2+} chelation alone did not lead to erbB2 phosphorylation (Fig. 2h). As expected, phosphorylation of erbB3 occurred under the same conditions (Fig. 2j). These functional data complement the biochemical and immunocytochemical data and indicate that tight junctions segregate ASL heregulin- α from its receptors erbB2–4 on the basolateral surface.

This physical arrangement creates a ligand–receptor system poised for activation whenever epithelial integrity is compromised. To test this concept, we mechanically injured epithelia. Scanning electron microscopy showed that the device scraped off a ring of

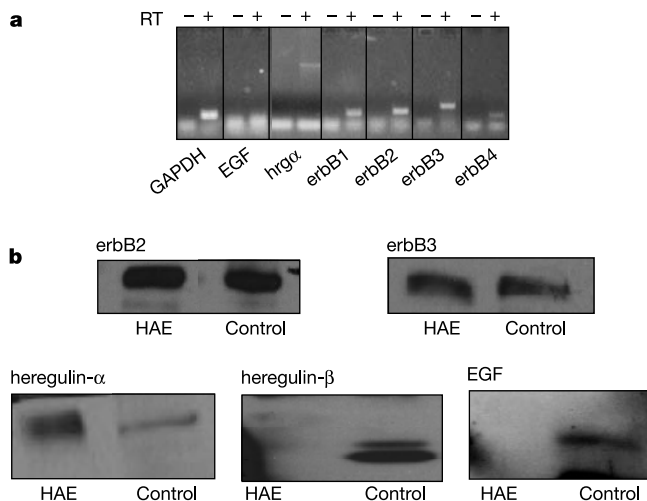


Figure 1 RT–PCR and western blots from primary cultures of human airway epithelia. **a**, RNA was subjected to PCR in the absence (–) or presence (+) of reverse transcriptase (RT). Plasmid containing the appropriate complementary DNA served as positive control. EGF, epidermal growth factor; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; hrg- α , heregulin- α . **b**, Human airway epithelial (HAE) cell lysates and positive control lysates were separated by SDS–PAGE and subjected to western blot analysis. Positive controls were: erythroleukaemia cell line K562 stably transfected with full length cDNAs for erbB2 and erbB3; MDA-MB231 breast cancer cell line, which endogenously expresses heregulin- α ; lung cancer cell line 522 which endogenously expresses heregulin- β ; and recombinant EGF.

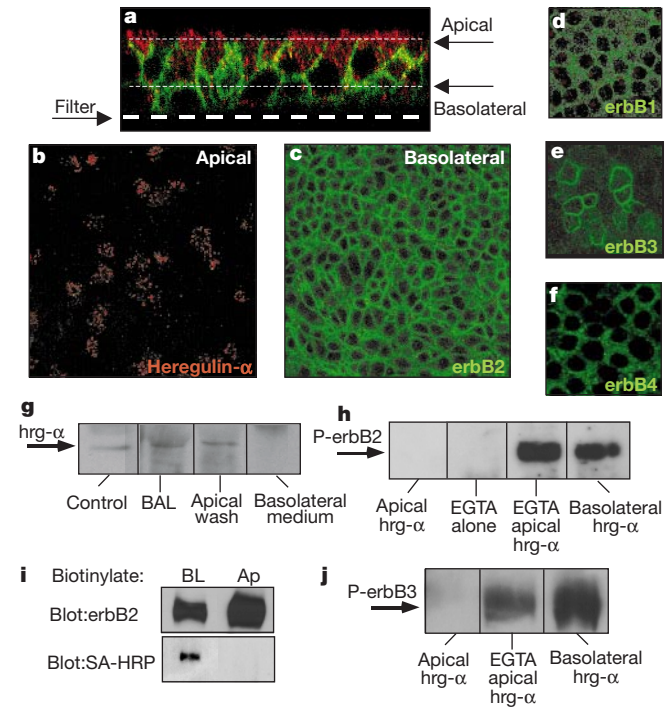


Figure 2 Immunostaining of erbB1–4 and heregulin- α in airway epithelia. **a**, Confocal image in x – z plane: erbB2 (green) and heregulin- α (red). **b**, **c**, *En face* images at positions indicated in **a**; apical (**b**) and basolateral (**c**) membranes. **d**–**f**, *En face* images of basolateral membrane immunostained for erbB1, 3 and 4. **g**, Western blot analysis of heregulin- α in airway surface liquid (ASL) *in vitro* and *in vivo*. Control, positive lysate from MDA-MB231 cells; BAL, bronchoalveolar lavage liquid; apical wash, collected ASL; and basolateral medium, medium collected from epithelia. **h**, Western blot analysis of phosphorylated erbB2 following basolateral addition of heregulin- α (30 nM) or apical addition following Ca^{2+} chelation. Apical application of heregulin- α alone or EGTA treatment alone are negative. **i**, Western blot analysis of apical (Ap) or basolateral (BL) surface biotinylated proteins immunoprecipitated for erbB2. **j**, Western blot analysis of phosphorylated erbB3 following basolateral addition of heregulin- α or its apical addition following Ca^{2+} chelation. Apical application of heregulin- α alone was negative.

cells, leaving only the permeable support at the site (Fig. 3a); cells bordering the wound remained intact (inset, high magnification). We examined the role of heregulin- α and erbB2-4 in the repair process by measuring transepithelial electrical conductance (G_t). Intact epithelia maintained a low G_t over time (Fig. 3c). Following the injury, G_t increased abruptly, remained elevated for 12–18 h, and then slowly recovered to basal values by ~24 h. Application of recombinant heregulin- α accelerated repair (Fig. 3d). Conversely, application of an antibody that blocks the erbB2 receptor (herceptin) significantly slowed G_t recovery. Western blot analysis confirmed erbB2 phosphorylation 5 min following wounding (Fig. 3b). Thus, heregulin- α and erbB2-4 contributed to repair of human airway epithelia following a mechanical wound. These data suggest that the injury allowed ASL heregulin- α to activate its basolateral receptor. We also tested the contribution of heregulin- α and erbB receptors to wound repair using human tracheal explants bathed with medium containing ASL. Following a mechanical injury, the wound healed over in ~18 h (Fig. 3f). Blocking ligand with a neutralizing heregulin- α antibody or blocking erbB2 with herceptin inhibited wound repair (Fig. 3e, f).

To further test receptor activation, we used immunocytochemistry with an antibody specific for the phosphorylated form of erbB2. Within 1 min after wounding, receptors were activated (Fig. 4a, b). Antibody staining was present only at the edge of the wound, consistent with a localized effect of heregulin- α derived from the overlying ASL. In contrast, we saw diffuse phospho-erbB2 staining when we applied heregulin- α basolaterally or when we applied it apically and disrupted the tight junctions by Ca^{2+} chelation (Fig. 4c, d). Calcium chelation in the absence of apical heregulin- α did not lead to erbB2 phosphorylation (Fig. 4e).

Following mechanical injury, ASL growth factors other than heregulin- α might also activate erbB2. To test this possibility, we added a neutralizing anti-heregulin- α antibody to the apical surface before mechanical injury. Neutralizing ASL heregulin- α eliminated erbB2 activation at the edge of the wound (Fig. 4g), whereas phosphorylated erbB2 was evident in control epithelia and epithelia treated with non-immune antibody (Fig. 4f, h). These results support our biochemical and immunolocalization data. They also emphasize the importance of ASL heregulin- α in activating basolateral erbB growth factor receptors. Of course, these data do not exclude a role for ligands such as EGF that can be produced in the basolateral compartment by fibroblasts.

We also tested segregation of ASL ligand from basolateral receptor using intact rat trachea. We isolated tracheae, removed ASL by washing the lumen, closed both ends, and placed them in an EGTA-containing solution to open the tight junctions. When transepithelial conductances increased, we isolated the epithelial cells, immunoprecipitated erbB2, and probed the blot with an anti-phosphotyrosine antibody. erbB2 was phosphorylated when ASL was present, but phosphorylation was minimal when ASL had been washed away (Fig. 4i). We also mechanically wounded the trachea in anaesthetized rats and then immediately removed epithelial cells. Figure 4j shows that wounding led to erbB2 phosphorylation.

The physical segregation of a ligand to the apical solution and its receptor to the basolateral membrane provides a simple system that is primed for activation the instant epithelial integrity is compromised (see Supplementary Information). Such a design could be central to rapid restoration of barrier function and protection of the organism following a mechanical or toxic wound that disrupts the epithelium. Our findings also suggest models in which other ASL factors might activate specific basolateral receptors, and conversely, basolateral ligands might access apical receptors when tight-junction integrity is compromised. Finally, this model could apply to any epithelial barrier.

This ligand–receptor arrangement may also be critical in other physiological and pathophysiological situations (see Supplementary Information). For example, during development unpolarized cells

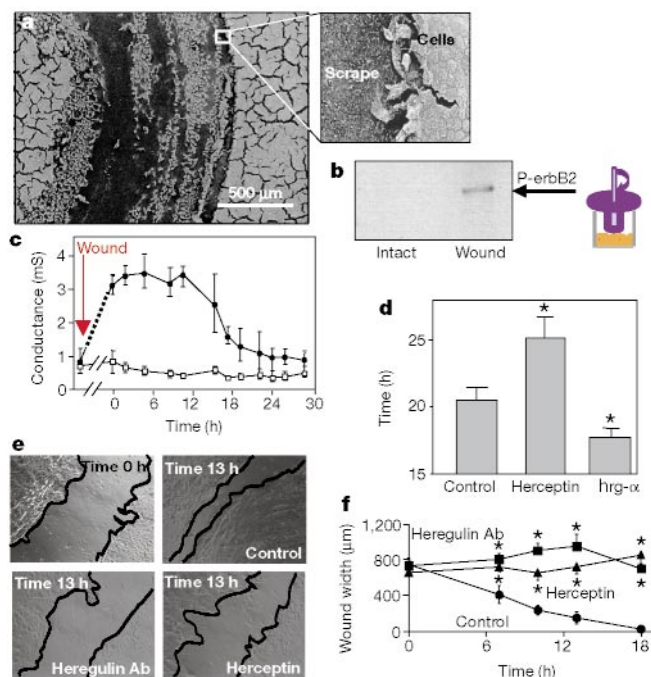


Figure 3 Involvement of ASL heregulin- α and erbB2 following mechanical wounding and repair of airway epithelia. **a**, Scanning electron micrograph of wounded airway epithelium; wounding device shown schematically at right. Inset, higher-magnification view at wound edge. **b**, Western blot analysis of phosphorylated erbB2 following injury. **c**, Time course of transepithelial electrical conductance (G_t) in wounded (filled circles) and intact epithelia (open squares) ($n = 6$). **d**, Time for 50% recovery of G_t following injury. Data are mean $\pm$ s.e.m. ($n = 6$) for control (wound only) epithelia, wounded epithelia treated with either anti-erbB2 antibody (3.2 μM herceptin) or recombinant heregulin- α protein (1.7 μM). Asterisks indicate $P < 0.02$ compared to control. **e**, Phase-contrast microscopy images of injured human tracheal explants; wound edges highlighted for clarity. Wound width at the start of experiment is shown at time 0 h. At 13 h post-injury, control wound width decreased while that of explants treated with neutralizing heregulin- α antibody (heregulin Ab) or herceptin remained unchanged. **f**, Wound width as function of time. Data are mean $\pm$ s.e.m. ($n = 8$). Asterisks indicate $P < 0.01$ compared to controls.

might secrete ligands that target receptors on their own surface or that of adjacent cells. Differentiation and the development of tight junctions could segregate the ligand from its receptor, thereby silencing a signalling cascade. To test this idea, we studied airway cells 4 h after seeding, before they polarized and formed an epithelium (Fig. 4k–m). These cells did not segregate heregulin- α and erbB2—they were expressed over the cell surface—and the receptor was phosphorylated. In this experiment, the immunostained heregulin- α probably represents the unprocessed, cell surface form, which has been demonstrated to bind and activate erbB receptors¹⁰. Once tight junctions formed, ligand and receptor were segregated and receptor activation was silenced (Fig. 2h, and area away from the wound in Fig. 4b, f, h). This process could be important in mucosal cancers, where a loss of epithelial polarity, mislocalization of a growth factor receptor, or altered polarized sorting could continuously expose receptors to ligand, thereby stimulating proliferation^{6,11–13}.

This model also predicts that increased paracellular permeability, without complete epithelial disruption, would allow ligand to access its receptor. Consistent with this, Ca^{2+} chelation increased paracellular permeability, allowing ASL heregulin- α to access its receptor and phosphorylate erbB2 (Fig. 4d, i). Several airway diseases, including smoking-associated bronchitis, cystic fibrosis, and

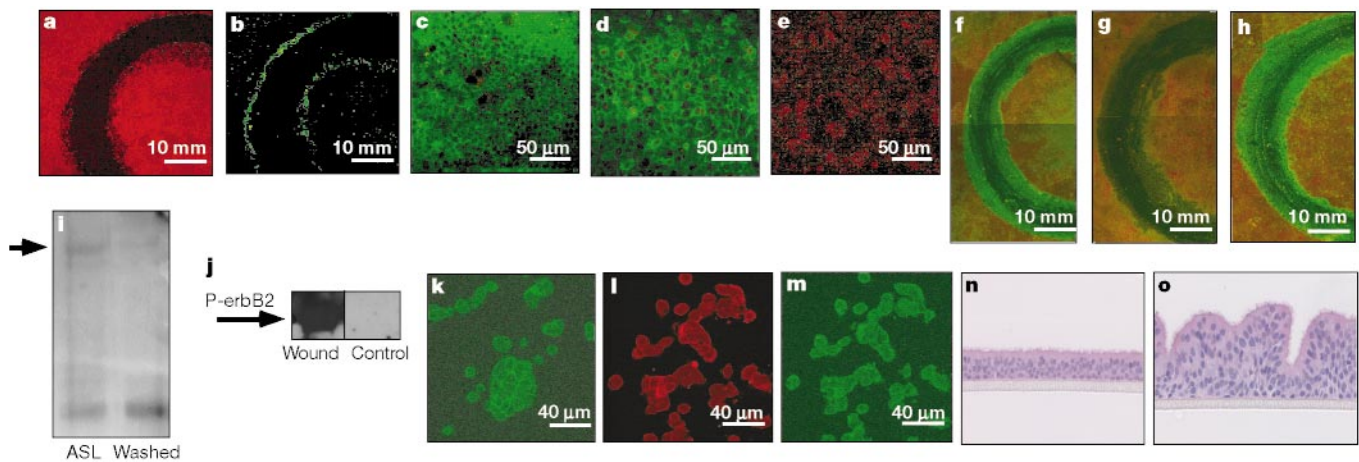


Figure 4 Immunostaining of phosphorylated erbB2. **a**, Ethidium bromide staining (red) shows absence of cells in wounded region. **b**, Immunostaining for phosphorylated erbB2 (green) indicates receptor activation at wound edge. **c–e**, Immunostaining for phosphorylated erbB2 following basolateral heregulin- α application (**c**), apical heregulin- α after Ca^{2+} chelation (**d**), or in absence of heregulin- α (**e**). Cells were counterstained with ethidium bromide (red). **f–h**, Mechanically injured epithelia immunostained for phosphorylated erbB2 (green), with red ethidium bromide counter-staining. **f**, Control conditions. **g**, Apical application (in 2 μl) of neutralizing anti-heregulin- α antibody before

injury. **h**, Apical application of non-immune antibody (2 μl) before injury. **i**, Western blot analysis for phosphorylated erbB2 (arrow) from freshly isolated rat tracheae treated with EGTA. ASL was present or removed by washing before treatment. **j**, Western blot analysis for phosphorylated erbB2 from rat tracheae mechanically injured *in vivo*. **k–m**, Unpolarized epithelia immunostained for erbB2 (**k**), heregulin- α (**l**), and phosphorylated erbB2 (**m**). **n, o**, Haematoxylin- and eosin-stained 1- μm sections from control airway epithelia (**n**) and epithelia treated with basolateral heregulin- α (30 nM) for 10 d (**o**). Images are representative of four studies.

asthma manifest a chronic increase in epithelial permeability and epithelial remodelling^{14–17}. To test the physiological consequences of chronic receptor stimulation with one ASL component, we applied heregulin- α to the basolateral medium for 10 d and found a thickening of the epithelium (Fig. 4n, o), indicating epithelial remodelling. We speculate that increased epithelial permeability and the resulting chronic receptor activation contribute to the remodelling that is a hallmark of those diseases. □

Methods

Cell culture

Human airway epithelia from lungs removed for organ donation were isolated and cultured at the air–liquid interface as described previously¹⁸. Epithelia were studied at least 2 weeks after seeding, when they had differentiated and developed a ciliated apical surface.

Immunocytochemistry and microscopy

Epithelia were immunostained as previously described¹⁹. Cultures were counter-stained with ethidium bromide, mounted on glass slides with Vectashield mounting media (Vector Labs) and studied using confocal microscopy (MRC-1024; Bio-Rad Laboratories).

Following prolonged heregulin- α treatment cultures were fixed, paraffin embedded, sections stained with haematoxylin and eosin, and analysed with an Olympus BX-51 using a spot RT slider CCD camera (Diagnostic Instruments). Following wounding, samples were processed for scanning electron microscopy as previously described¹⁹.

Antibodies

Unless otherwise noted, antibodies were used at 1:100 for immunostaining and 1:1,000 for immunoprecipitation and western blot. We used the following antibodies: anti-phosphorylated erbB2 (Y1248, Upstate Biotechnology); anti-erbB2 (Ab-1, Oncogene); anti-erbB2 (20-C, neu 9G6, Santa Cruz Biotechnology); anti-heregulin- α and β antibodies (Santa Cruz Biotechnology); anti-erbB3 (RTJ.1) and anti-EGFR (Pharmingen); anti-erbB2 (DAKO Corporation) at 1:500 for immunoprecipitation and western blots; and anti-phosphorylated erbB2 (PN2A, NeoMarkers) at 2 $\mu\text{g ml}^{-1}$ for western blots. The erbB2 blocking antibody (herceptin, Genentech) was used at 3.2 μM in primary cultures and at 1 mM in human tracheal explants. Heregulin- α neutralizing antibody (R&D Systems) was used at 100 $\mu\text{g ml}^{-1}$ for cultures, 0.3 μM for explants.

Recombinant protein

Recombinant heregulin- α (NeoMarkers or R&D Systems) was used at 1.7 μM for injured cultures, and 30 nM or 12 nM for basolateral treatment and for EGTA-treated cultures.

Immunoprecipitations and western blots

Lysate made by incubation in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM

EDTA, 2 mM Na_3VO_4 , 10 mM sodium pyrophosphate, 100 mM NaF, phenylmethylsulphonyl fluoride (PMSF), leupeptin, pepstatin, aprotinin) with 1% Triton X-100 was homogenized, spun, the soluble fraction immunoprecipitated and complexed with protein G Sepharose (Pierce). Protein was eluted by incubation in sample buffer (4% SDS, 100 mM dithiothreitol (DTT), 20% glycerol, 0.005% bromophenol blue, 0.065 M Tris, pH 6.8), boiled, separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and transferred to polyvinylidene fluoride membrane (Immobilon-P, Millipore Corporation). Membranes were blocked in 5% bovine serum albumin (BSA), washed and incubated with antibody. Incubation in horseradish peroxidase (HRP)–secondary antibody (1:10,000, Amersham Pharmacia) allowed for bound antibody detection with SuperSignal solution (Pierce) and exposure to film (X-OMAT AR, Kodak Scientific Imaging Film).

ASL was collected by washing the apical surface of multiple epithelia with phosphate buffered saline (PBS). Basolateral medium that had bathed the epithelia for 1 week was also collected. Heregulin- α was immunoprecipitated from 2 ml of apical washes and 10 ml medium as described above.

Wounding and measurement of transepithelial electrical conductance

A wounding device was designed to generate a consistent pressure and injury area on the epithelia (Fig. 3). When lowered onto cells and turned 360° four times, it scraped off a ring of cells generating a reproducible wound.

Following injury, transepithelial electrical conductance (G_t) was measured with an ohmmeter (World Precision Instruments). Eagle's modified essential medium was applied apically to measure G_t and then removed. As this removed apically applied herceptin or heregulin- α , these were re-applied following measurement.

Ca^{2+} chelation to disrupt tight junctions

Epithelial tight junctions were disrupted by apical treatment with 5 mM EGTA in water. G_t was monitored until it fell to background levels. EGTA solution was removed and recombinant heregulin- α protein (30 nM) apically applied.

RT-PCR

Total RNA was extracted using RNA-Stat 60 (Tel-test B) according to manufacturer's directions. DNA was removed by incubation with deoxyribonuclease I (Dnase I, Gibco BRL). RNA was reverse transcribed and aliquots of the reaction mix subjected to PCR (94°C for 2 min, 95°C for 30 s, 55°C for 30 s, 72°C for 1 min; 35 cycles) using the following specific primers: erbB1 forward: cagccacctgtgtcaacagc, reverse: aataaattcactgtttgttg; erbB2 forward: ggtgccgtggagaaacccgag, reverse: tcacactggcactgtccagacc; erbB3 forward: atgagggcgcaacgactctgt, reverse: gtgctgtgagttgtgttata; erbB4 forward: ggaaccacacgtgtgtgtg, reverse: caccacagtattcgtgtgtct; GAPDH forward: ggctacactgagcaccaggtgt, reverse: gtcttggaggccatgtgtggg; EGF forward: gggcacagagcaaggtgtgt, reverse: ctgagtcagctcattgtgtgtg; heregulin- α forward: gagatgtccgagcgcaagaagcgaga, reverse: tgaagacacatctgtcttcagttgaggc.

Study of rat tracheas

Male rats (Harlan Sprague Dawley) were euthanized with CO_2 , their tracheae immediately excised and the luminal surfaces washed with PBS (to remove ASL) before suturing both

ends. Control tracheae were not washed. Sutured tracheae were bathed in 50 mM EGTA; other portions of trachea were studied in Ussing chambers with continual G_t measurement. When G_t increased, tracheae were removed from the bath, epithelial cells isolated using an endocervical brush (Birchwood Laboratories, Inc.), placed in lysis buffer with 1% Triton X-100, immunoprecipitated with an anti-erbB2 antibody and blotted as described above.

For *in vivo* studies, male rats were anaesthetized with ketamine/xylazine, their necks dissected, tracheae exposed and injured with a bronchoscope brush. Tracheal epithelial cells were immediately isolated and analysed as described above. Rats were euthanized with CO_2 .

Human tracheal explant model

Human tracheae of lungs removed for organ donation were dissected into 2–3 mm² explants and grown on chambered coverglass (Nalge Nunc International) coated with human placental acid soluble collagen type VI substrate (Sigma)¹⁸. After 1 week in culture, explants were fed medium containing ASL, mechanically injured with a yellow pipette tip (BioExpress) and maintained with neutralizing heregulin- α antibody (0.3 μ M) or heregulin (1 mM) added to the medium. Wounds were examined with a Nikon Phase Contrast inverted microscope and photographed with a Spot RT camera (Diagnostic Instruments, Inc). The width of each wound was measured by an observer, blind to the experimental design, at 10 sites in each image.

Received 26 November 2002; accepted 17 January 2003; doi:10.1038/nature01440.

- Aida, S., Tamai, S., Sekiguchi, S. & Shimizu, N. Distribution of epidermal growth factor and epidermal growth factor receptor in human lung: immunohistochemical and immunoelectron-microscopic studies. *Respiration* **61**, 161–166 (1994).
- al Moustafa, A. E., Alaoui-Jamali, M., Paterson, J. & O'Connor-McCourt, M. Expression of P185erbB-2, P160erbB-3, P180erbB-4, and heregulin α in human normal bronchial epithelial and lung cancer cell lines. *Anticancer Res.* **19**, 481–486 (1999).
- Polosa, R. *et al.* Expression of c-erbB receptors and ligands in human bronchial mucosa. *Am. J. Respir. Cell Mol. Biol.* **20**, 914–923 (1999).
- van der Geer, P., Hunter, T. & Lindberg, R. A. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* **10**, 251–337 (1994).
- Walker, R. A. The erbB/HER type 1 tyrosine kinase receptor family. *J. Pathol.* **185**, 234–235 (1998).
- Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nature Rev. Mol. Cell Biol.* **2**, 127–137 (2001).
- Sweeney, C. *et al.* Growth factor-specific signaling pathway stimulation and gene expression mediated by ErbB receptors. *J. Biol. Chem.* **276**, 22685–22698 (2001).
- Wang, G. *et al.* Influence of cell polarity on retrovirus-mediated gene transfer to differentiated human airway epithelia. *J. Virol.* **72**, 9818–9826 (1998).
- Dempsey, P. J., Meise, K. S., Yoshitake, Y., Nishikawa, K. & Coffey, R. J. Apical enrichment of human EGF precursor in Madin-Darby canine kidney cells involves preferential basolateral ectodomain cleavage sensitive to a metalloprotease inhibitor. *J. Cell Biol.* **138**, 747–758 (1997).
- Aguilar, Z. & Slamon, D. J. The transmembrane heregulin precursor is functionally active. *J. Biol. Chem.* **276**, 44099–44107 (2001).
- Campiglio, M., Ali, S., Knyazev, P. G. & Ullrich, A. Characteristics of EGFR family-mediated HRG signals in human ovarian cancer. *J. Cell. Biochem.* **73**, 522–532 (1999).
- Mostov, K. E., Verges, M. & Altschuler, Y. Membrane traffic in polarized epithelial cells. *Curr. Opin. Cell Biol.* **12**, 483–490 (2000).
- Wodarz, A. Tumor suppressors: linking cell polarity and growth control. *Curr. Biol.* **10**, R624–R626 (2000).
- Kennedy, S. M., Elwood, R. K., Wiggs, B. J., Pare, P. D. & Hogg, J. C. Increased airway mucosal permeability of smokers. Relationship to airway reactivity. *Am. Rev. Respir. Dis.* **129**, 143–148 (1984).
- Sobonya, R. E. & Taussig, L. M. Quantitative aspects of lung pathology in cystic fibrosis. *Am. Rev. Respir. Dis.* **134**, 290–295 (1986).
- Godfrey, R. W. Human airway epithelial tight junctions. *Microsc. Res. Tech.* **38**, 488–499 (1997).
- Goto, Y. *et al.* Dislocation of E-cadherin in the airway epithelium during an antigen-induced asthmatic response. *Am. J. Respir. Cell Mol. Biol.* **23**, 712–718 (2000).
- Karp, P. H. *et al.* in *Epithelial Cell Culture Protocols* (ed. Wise, C.) 115–137 (Humana, Totowa, New Jersey, 2002).
- Walters, R. W. *et al.* Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. *Cell* **110**, 789–799 (2002).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank P. Karp, P. Weber, J. Launspach, T. Nesselhauf, L. Panko, T. Mayhew and D. Vermeer for technical assistance. We appreciate the help of the Iowa Statewide Organ Procurement Organization. We also thank the University of Iowa *In Vitro* Models and Cell Culture Core (supported in part by the National Heart, Lung and Blood Institute (NHLBI), the Cystic Fibrosis Foundation, and the National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK)), and the University of Iowa Central Microscopy Research Facility. This work was supported by the NHLBI, the CFF, and the Howard Hughes Medical Institute. P.D.V. was supported by a K01 award from the NIDDK. M.J.W. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.Z.
(e-mail: joseph-zabner@uiowa.edu) or M.J.W. (e-mail: michael-welsh@uiowa.edu).

The Par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein Lgl

Jörg Betschinger, Karl Mechtler & Juergen A. Knoblich

Research Institute of Molecular Pathology, Dr Bohr Gasse 7, 1030 Vienna, Austria

To generate different cell types, some cells can segregate protein determinants into one of their two daughter cells during mitosis. In *Drosophila* neuroblasts, the Par protein complex localizes apically^{1–5} and directs localization of the cell fate determinants Prospero^{6–8} and Numb⁹ and the adaptor proteins Miranda^{10,11} and Pon¹² to the basal cell cortex, to ensure their segregation into the basal daughter cell. The Par protein complex has a conserved function in establishing cell polarity¹³ but how it directs proteins to the opposite side is unknown. We show here that a principal function of this complex is to phosphorylate the cytoskeletal protein Lethal (2) giant larvae (Lgl; also known as L(2)gl). Phosphorylation by *Drosophila* atypical protein kinase C (aPKC), a member of the Par protein complex, releases Lgl from its association with membranes and the actin cytoskeleton. Genetic and biochemical experiments show that Lgl phosphorylation prevents the localization of cell fate determinants to the apical cell cortex. Lgl promotes cortical localization of Miranda^{14,15}, and we propose that phosphorylation of Lgl by aPKC at the apical neuroblast cortex restricts Lgl activity and Miranda localization to the opposite, basal side of the cell.

The *Drosophila* Par protein complex consists of aPKC⁵ and the conserved PDZ domain proteins Bazooka (Par-3)^{1–3} and Par-6 (ref. 4). Apical Par proteins direct localization of determinants to the opposite, basal cortex (Fig. 1a). To identify additional complex members that might be required for basal protein localization, proteins that immunoprecipitated together with Par-6 from *Drosophila* embryos were microsequenced by mass spectroscopy (Fig. 1b). Two protein bands with relative molecular masses of approximately 66,000 and 120,000 (M_r 66K and 120K, respectively) were identified as aPKC and Lgl, respectively. Lgl was not known to be part of the Par protein complex, and to verify its interaction we generated anti-Lgl antibodies and used them in immunoprecipitations. Anti-Lgl specifically co-immunoprecipitates aPKC and Par-6 (Fig. 1c). Conversely, both aPKC and Lgl can be detected in western blots of Par-6 immunoprecipitations (Fig. 1d). When the Par complex is affinity purified from embryos on glutathione S-transferase (GST)–Par-6 beads, both aPKC and Lgl can be detected (Fig. 1e). When the complex is purified on maltose-binding protein (MBP)–Bazooka beads, however, Par-6 and aPKC are still detected but Lgl is not present (Fig. 1f), indicating that Bazooka and Lgl might be mutually exclusive members of the aPKC–Par-6 complex.

To determine how Lgl is recruited into the Par protein complex, we performed *in vitro* binding experiments (see Fig. 1g). Lgl contains four WD40 repeats (Fig. 1g, red). Par-6 has a conserved amino-terminal domain (green), a CRIB domain (blue) and a PDZ domain (red). *In vitro* translated Lgl and aPKC bind GST–Par-6 but not GST alone (Fig. 1h). Par-6 lacking 121 amino acids of the conserved N terminus no longer binds aPKC (Fig. 1h, Δ 1). In contrast, the carboxy-terminal PDZ domain of Par-6 (residues 146–280) is required (Fig. 1h, Δ 3) and sufficient (Fig. 1h, PDZ) for binding to Lgl. A deletion analysis of Lgl (Fig. 1i, j) shows that both the N terminus and the central part (Fig. 1j) but not the C terminus of the protein can bind Par-6.

Lgl is required for asymmetric cell division^{14,15}; in *lgl* mutant

neuroblasts, Par proteins localize normally to the apical cortex but fail to direct basal localization of Miranda and Prospero proteins. Lgl itself is not asymmetrically distributed in mitotic neuroblasts. It is found in the cytoplasm and uniformly around the cortex. The different localizations of Lgl and the Par proteins suggest that a modified version of Lgl may be present in the Par protein complex. Lgl is phosphorylated by an associated kinase¹⁶. To test whether this kinase is contained in the Par protein complex, the complex was immunopurified using anti-Par-6 antibodies and incubated with radiolabelled ATP (Fig. 2a). In addition to a band at 49K of

unknown identity, we detected a strong radiolabelled band at the molecular mass of Lgl. We detected no other protein of similar molecular mass in the Par protein complex by silver staining (Fig. 1a), suggesting that Lgl is phosphorylated by aPKC in this complex. To determine the phosphorylation site, we used PKC ζ , the mammalian homologue of *Drosophila* aPKC, to phosphorylate bacterially produced Lgl (Fig. 2c). Both the full-length protein (Lgl) and the central region (Lgl-m, residues 402–802)—but neither the N- (Lgl-N, residues 1–403) nor the C terminus (Lgl-C, residues 800–1161)—are phosphorylated by PKC ζ *in vitro*. Of eight serine residues within the central fragment that match the consensus for PKC phosphorylation (S-X(R/K)), three are conserved between flies and humans (Fig. 2b). When any two of these sites are mutated to alanine, Lgl is still phosphorylated. Mutation of all three sites, however, completely abolishes phosphorylation by PKC ζ (Fig. 2d).

To test whether Lgl is phosphorylated by aPKC *in vivo*, we generated antibodies against a peptide corresponding to residues 655–666 of Lgl with phosphorylated serine at position 660 (anti-pLgl). The antibodies are specific and recognize bacterially expressed Lgl-m only after phosphorylation by PKC ζ (Fig. 3a). Lgl immunoprecipitated from *Drosophila* embryos is detected by the phospho-specific antibody (Fig. 3b), indicating that it is phosphorylated on serine 660. Lgl immunoprecipitating together with Par-6 is also recognized by the phospho-specific antibody (Fig. 3b, Par-6). The pLgl signal increases after addition of ATP (Fig. 3b, Par-6 plus ATP), indicating that serine 660 can be phosphorylated by a protein kinase that is present in the complex. To test whether this kinase is aPKC, we used *Drosophila* Schneider (S2) cells where proteins can easily be depleted by RNA-mediated interference (RNAi) (Fig. 3c). S2 cells express Lgl, aPKC and Par-6,

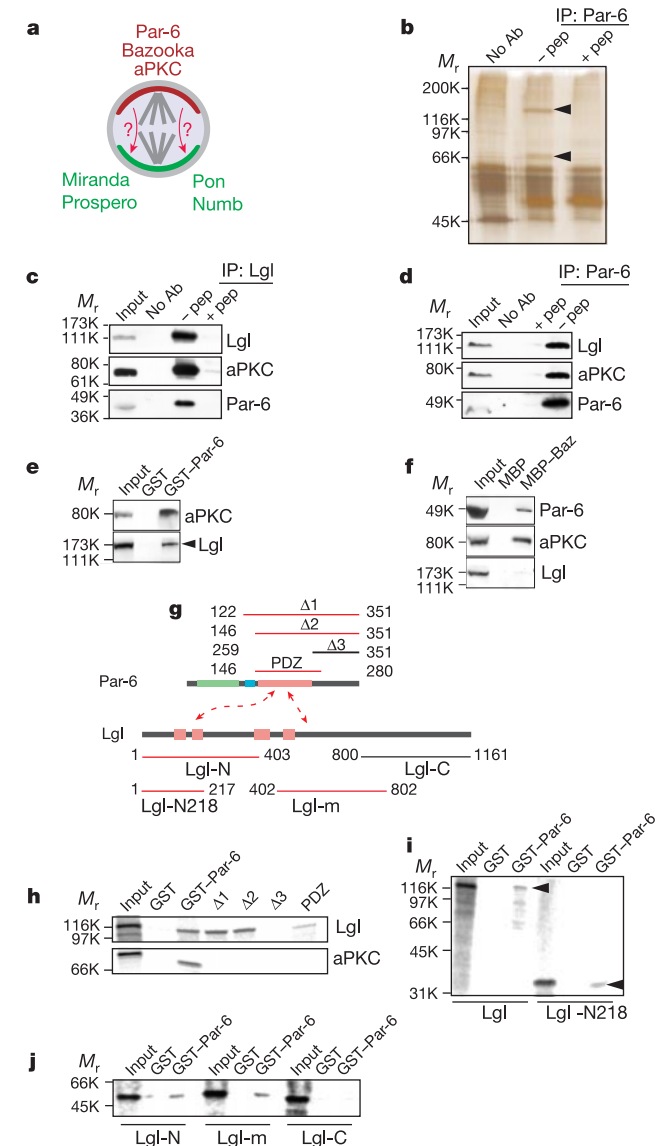


Figure 1 Lgl is in the Par protein complex. **a**, Asymmetric protein localizations in neuroblasts. **b**, Silver stain of immunoprecipitations from 3–7-h-old *Drosophila* embryos using protein A beads (no antibody (Ab)), anti-Par-6 or anti-Par-6 blocked by peptide antigen. Arrowheads indicate sequenced bands. Par-6 runs with immunoglobulin- γ (IgG) and is not resolved. **c**, **d**, Immunoblots of anti-Lgl and anti-Par-6 immunoprecipitations. Input is one-tenth of the extract used, controls are beads (no Ab) and peptide-blocked antibody (+pep). **e**, **f**, Immunoblots of GST–Par-6 (**e**) or MBP–Bazooka (**f**) beads incubated with embryonic extract. **g**, Par-6 (green, N-terminal conserved domain; blue, CRIB domain; pink, PDZ domain) and Lgl (pink, WD40 repeats) constructs used (red, binding; black, no binding). **h**, Binding of *in vitro* translated Lgl and aPKC to GST fusion proteins ($\Delta 1$, GST–Par-6 $\Delta 1$; $\Delta 2$, GST–Par-6 $\Delta 2$; $\Delta 3$, GST–Par-6 $\Delta 3$; PDZ, GST–Par-6-PDZ). Input in **h–j** is one-tenth of the amount used in the experiment. **i**, **j**, *in vitro* binding of Lgl deletions to GST and GST–Par-6.

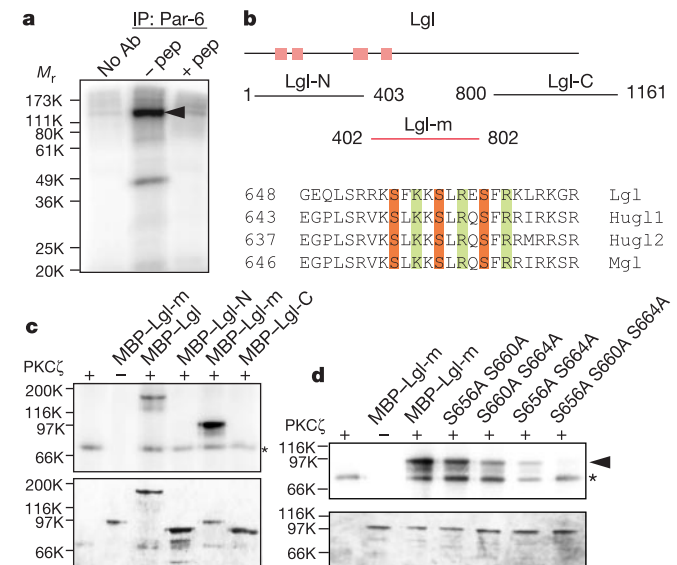


Figure 2 aPKC phosphorylates Lgl on three conserved serine residues *in vitro*.

a, Autoradiograph of anti-Par-6 immunoprecipitate incubated with [γ -³²P]-labelled ATP. Protein A beads (no antibody) or peptide-blocked antibody (+pep) are controls. The arrowhead marks a phosphorylated 120K band. **b**, Summary of Lgl constructs used in kinase assays (pink, WD40 repeats). Numbers indicate first and last amino acid in each construct. Three conserved serine residues (orange) match the PKC consensus (S-X(R/K), green) in the *Drosophila*, mouse (Mgl) and both of the human (Hugl-1 and Hugl-2) homologues. **c**, Human PKC ζ (the aPKC homologue) phosphorylates MBP–Lgl and MBP–Lgl-m but not MBP–Lgl-N and MBP–Lgl-C. Controls without substrate and without kinase are shown (top, autoradiograph; bottom, Coomassie stain). The asterisk marks autophosphorylated PKC ζ . **d**, PKC ζ phosphorylates MBP–Lgl-m but not a mutant in which serine residues 656, 660 and 664 are mutated to alanine. All double mutant combinations are phosphorylated. Controls and panels as in **c**.

and Lgl immunoprecipitated from S2 cells is recognized by the phospho-specific antibody (Fig. 3c, control). On depletion of aPKC, however, Lgl phosphorylation is strongly reduced (Fig. 3c, aPKC^{RNAi}). Similar results were obtained using an antibody that recognizes phosphorylated serine 664 (data not shown). Thus, Lgl is an *in vivo* substrate for aPKC.

We used S2 cells to determine whether phosphorylation by aPKC affects the properties of the Lgl protein. Endogenous Lgl in S2 cells is hard to detect by immunofluorescence (data not shown). After transfection, both wild-type Lgl and a mutant form, in which all three aPKC phosphorylation sites are mutated to alanine (Lgl-3A), are localized at the cell cortex. Co-transfection of aPKC releases Lgl but not Lgl-3A from the cell cortex (Fig. 3d), indicating that phosphorylation changes the properties of the Lgl protein. To test how phosphorylation affects endogenous Lgl, we fractionated embryo extracts by high-speed centrifugation (Fig. 3e). α -Tubulin is unpolymerized under the conditions used and marks the cytosol, whereas syntaxin and myosin II heavy chain mark the membrane and cytoskeletal fraction, respectively. Lgl exists in a soluble, cytoplasmic form and in an insoluble form that is thought to be

associated with membranes and the actin cytoskeleton¹⁷. Similar to aPKC and Par-6, it is therefore present both in the pellet and supernatant. Phosphorylated Lgl, however, is mostly cytosolic, indicating that it is neither membrane nor actin associated. Anti-Lgl immunoprecipitations from the cytosolic fraction and the re-extracted pellet (Fig. 3f) show that the cytosolic protein is no longer bound to Par-6 and aPKC. Thus, phosphorylation releases Lgl from the Par complex and from actin and membranes.

As the proposed functions for Lgl involve binding to myosin or membrane proteins^{18–21}, phosphorylation by aPKC might actually inactivate the protein at the apical cortex of neuroblasts. Anti-pLgl

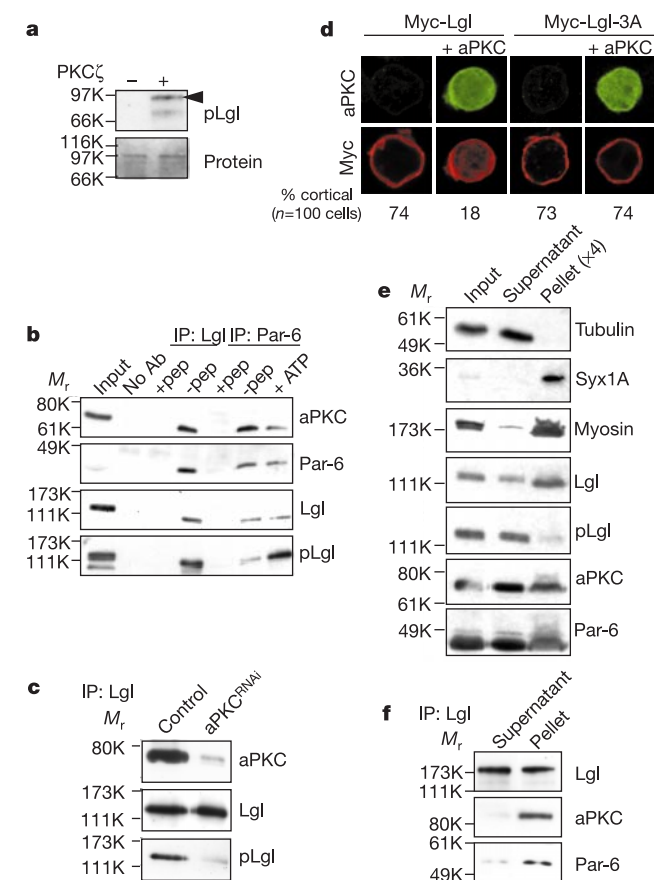


Figure 3 aPKC phosphorylates Lgl *in vivo*. **a**, Anti-pLgl immunoblot of bacterially produced Lgl-m incubated with (+) or without (-) PKC ζ . **b**, Immunoblot of anti-Lgl and anti-Par-6 immunoprecipitates. Incubation with ATP before immunoblotting (+ATP) enhances the pLgl signal. Protein A beads (no Ab) and peptide blocks (+pep) are controls; input is one-tenth of extract used. Par-6 runs with yolk proteins and is not well resolved in input. **c**, Anti-Lgl immunoprecipitates from S2 cells before and after depletion of aPKC by RNAi. **d**, Co-transfection of aPKC releases Myc-Lgl but not Myc-Lgl-3A from the cortex in S2 cells. One hundred Myc- (and aPKC)-positive cells were analysed for statistics. **e**, 100,000g (1 h) supernatant and pellet (see Methods) probed for α -tubulin, syntaxin-1A, myosin II heavy chain, Lgl, pLgl, aPKC and Par-6. Input is total extract and pellet is four times more than supernatant and input. **f**, Immunoblot of anti-Lgl immunoprecipitation from supernatant or re-extracted pellet.

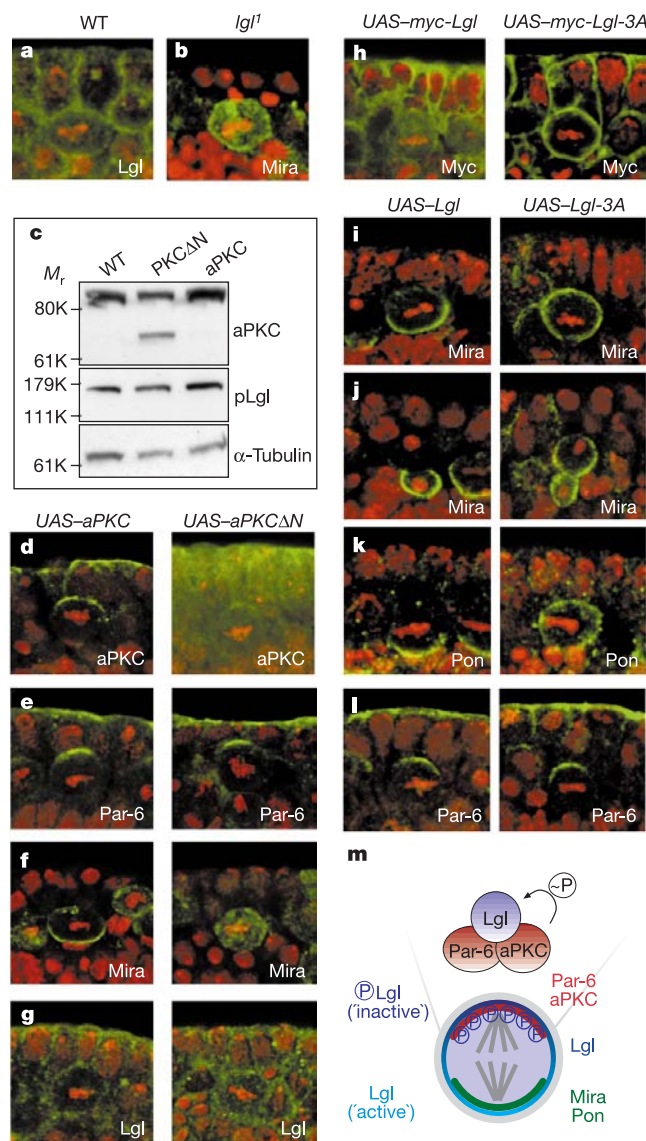


Figure 4 N-terminally truncated aPKC and non-phosphorylatable Lgl (Lgl-3A) inhibit asymmetric protein localization. **a**, Endogenous Lgl (anti-Lgl, green) is both cortical and cytoplasmic. DNA is stained red and the apical position is at the top in all panels. **b**, Miranda is mostly in the cytoplasm in *lgl*¹ germline clones. **c**, pLgl levels are not increased in UAS-aPKC Δ N or UAS-aPKC embryos. Transgenes are expressed from maternal Gal4 V32 in **c–i**. **d–g**, UAS-aPKC (left panels) or UAS-aPKC Δ N (right panels) embryos stained for aPKC (**d**), Par-6 (**e**), Miranda (**f**) or Lgl (**g**). **h**, UAS-myc-Lgl (left) or UAS-myc-Lgl-3A (right) embryos stained for Myc tag. **i–l**, UAS-Lgl (untagged) (left panels) and UAS-Lgl-3A (untagged) (right panels) embryos stained for Miranda (**i**, **j**), Pon (**k**) or Par-6 (**l**). Fly strains used in **i–l** express comparable amounts of Lgl protein (data not shown). **l**, A model in which apical phosphorylation of Lgl by the Par protein complex directs Miranda to the basal cell cortex.

is not specific in immunofluorescence and does not allow us to test the subcellular distribution of pLgl. We therefore expressed an N-terminally truncated form of aPKC (aPKC Δ N²²) that lacks the Par-6-binding domain in *Drosophila* embryos. In contrast to over-expressed wild-type aPKC (Fig. 4d), the truncated form is no longer restricted to the apical cortex but is also found basally and in the cytoplasm (Fig. 4d, right). Although it does not significantly increase overall pLgl levels (Fig. 4c), it is expected to phosphorylate Lgl also on the basal side. Neither aPKC nor aPKC Δ N affect Par-6 localization (Fig. 4e). Although Miranda localization is unaffected by aPKC overexpression (Fig. 4f), it relocalizes into the cytoplasm in 100% ($n = 18$) of neuroblasts overexpressing aPKC Δ N (Fig. 4f, right). This phenotype is similar to *lgl* mutants (Fig. 4b), suggesting that phosphorylation of Lgl on the basal side inactivates the protein. However, aPKC Δ N expression does not completely release Lgl from the cell cortex (Fig. 4g), thus interaction with another protein might retain its cortical localization. Our results suggest that Lgl is inactivated by local aPKC activity on the apical side of the neuroblast cell cortex.

To assess further the role of Lgl phosphorylation *in vivo*, we expressed Myc-tagged Lgl-3A (Myc-Lgl-3A) in neuroblasts. Endogenous Lgl protein (Fig. 4a) and overexpressed Myc-tagged wild-type Lgl (Myc-Lgl, Fig. 4h) are found both in the cytoplasm and at the cell cortex. Myc-Lgl-3A, however, is exclusively cortical (Fig. 4h, right), indicating that phosphorylation is required (but not sufficient) to release Lgl from the cell cortex. Lgl is required for cortical localization of Miranda in neuroblasts^{14,15} (Fig. 4b). It is possible that Miranda is excluded from the apical cortex because Lgl is inactivated through aPKC-dependent phosphorylation on this side. To test this, we analysed Miranda distribution in neuroblasts expressing Lgl-3A. In neuroblasts overexpressing wild-type Lgl, Miranda is found at the basal cortex in metaphase (Fig. 4i, 100% of $n = 26$ neuroblasts) and segregates into the basal daughter cell (Fig. 4j, 100% of $n = 21$ neuroblasts). When equal amounts of Lgl-3A are expressed, Miranda is localized around the whole cortex in metaphase (Fig. 4i, right, 47% of $n = 53$ neuroblasts) and segregates into both daughter cells in telophase (Fig. 4j, right, 47% of $n = 45$ neuroblasts). Similar to *lgl* mutants, spindle orientation and apical localization of Par proteins are unaffected in neuroblasts expressing Lgl-3A (Fig. 4l). We conclude that Lgl phosphorylation is necessary to exclude Miranda from the apical cell cortex. Expression of Lgl-3A also affects the asymmetric localization of Pon (Fig. 4k), Numb and Prospero (data not shown); however, expression of Lgl-3A does not affect epithelial apical-basal polarity (Fig. 4l, right and data not shown), indicating that other substrates for aPKC exist.

Our results suggest that Par proteins inhibit Lgl at the apical cell cortex of neuroblasts (Fig. 4m). Lgl is required for cortical recruitment of Miranda^{14,15} and this is why Miranda is only found at the basal cell cortex. Lgl and non-muscle myosin II show negative genetic interaction^{14–16,19,21,23} but the precise role of myosin II in Miranda localization remains to be determined. In yeast cells double mutant for the Lgl homologues *sro7* and *sro77*, vesicles involved in exocytosis fail to fuse with the plasma membrane²⁰. Lgl can bind plasma membrane t-SNAREs^{18,20} and could establish vesicle docking sites at the cell cortex. In neuroblasts, apical Lgl phosphorylation could restrict such docking sites to the basal side. Miranda is not a transmembrane protein but vesicular transport of a membrane protein might establish a high-affinity binding site for Miranda at the basal cortex. Not all functions of the Par protein complex are affected by Lgl-3A expression, suggesting that other downstream targets exist. In *Caenorhabditis elegans* zygotes and *Drosophila* oocytes, Par proteins are required for restricting the localization of the protein kinase Par-1 to the opposite side of the cell¹³. Restricting the localization or activity of other proteins to the opposite cell cortex seems to be a general principle by which the Par protein complex influences cell polarity. □

Methods

Mass spectroscopy and protein binding assays

Immunoprecipitations were performed from *white* mutant embryos essentially as described² (washes were 3×5 min and 3×15 min). SDS gels (10%) of bound proteins were silver stained²⁴. Excised bands were reduced with dithiothreitol, carboxy methylated using iodoacetamide and digested with trypsin. Peptides were extracted with formic acid and separated by nano-high-performance liquid chromatography on a PepMap C18 reversed-phase column. The column eluate was applied online to a LQC ion trap nanospray mass spectrometer (Finnigan). Spectra were analysed using MASCOT software (Matrix Science). Identified peptides cover 46% and 47% of the full-length Lgl and aPKC sequence, respectively. *In vitro* binding assays were performed as described²⁴ using a buffer containing 10 mM Tris, pH 7.5, 25 mM NaCl, 5 mM MgCl₂, 0.1% NP40 and 1 mM DTT. Protein binding assays to MBP-Bazooka or GST-Par-6 beads (Fig. 1e, f) were essentially as described⁴.

Fractionation and kinase assays

For fractionation, 4–7-h-old embryos were lysed in 10 mM Tris, pH 8.0, 1 mM EDTA, 1 mM phenylmethyl sulphonyl fluoride, precleared (centrifugation for 10 min, 4,000g) and centrifuged for 1 h at 100,000g through a 250-mM sucrose cushion. Pellets were re-extracted in IP buffer² for immunoprecipitations. For IP kinase assays, Par-6 immunoprecipitates were incubated for 30 min (30 °C) in kinase buffer (25 mM Tris, pH 7.5, 25 mM NaCl, 5 mM MgCl₂, 0.5 mM EGTA, 1 mM DTT) containing 50 μ M [γ -³²P]-labelled ATP. *In vitro* kinase assays were as described²⁵ using human PKC ζ (Calbiochem) for 30 min.

Constructs, flies and antibodies

Par-6 GST fusions were generated in pGEX4T-1; MBP-Lgl and MBP-Bazooka fusions in pMAL-c2x (NEB). For *in vitro* translation, Lgl complementary DNA containing two N-terminal Myc tags and a β -globin leader was used. Deletions were created by polymerase chain reaction; point mutants were generated using the QuickChange kit (Stratagene). For transgenic flies, Lgl or Lgl-3A with or without nine C-terminal Myc tags and truncated (aPKC Δ N, residues 180–606 (ref. 22)) or full-length aPKC were cloned into pUAST²⁶ and expressed using maternal GAL4 V32 (ref. 27). *Lgl*¹ germline clones were generated as described¹⁴. Immunofluorescence and microscopy were as described²⁷. Antibodies were: rabbit anti-Lgl (against the peptide HEKTNNGDNKIGTPKTAPEESQF, affinity purified, 1:100 dilution), rabbit anti-pLgl (peptide KSFKK(pS)LRSEFRC, affinity purified, 1:100), rabbit anti-PKC ζ (Santa Cruz, 1:1,000), mouse anti-c-Myc (Santa Cruz, 1:100), mouse anti-Par-6 (generated against GST-Par-6 Δ 1, Fig. 1g), rabbit anti Par-6 (ref. 4) (1:300), mouse anti- α -tubulin (Amersham, 1:100), mouse anti-syntrophin 1A²⁸, rabbit anti-non-muscle myosin²⁹ (1:1,000), rabbit anti-Miranda¹¹ (1:1,000) and rabbit anti-Pon¹² (1:1,000).

Cell culture

S2 cells were propagated in Schneider's medium (Gibco) containing 10% FCS, 50 U ml⁻¹ penicillin and 50 μ g ml⁻¹ streptomycin, and stained essentially as described³. Constructs were cloned into pUC-hygMT (provided by C. Thummel) and transfected using Cellfectin (Invitrogen). RNAi was performed as described³⁰ for 5 days in 3 ml serum-free Schneider's medium using 6–10 μ g double-stranded RNA (about 10 nM) corresponding to nucleotides 1–700 of the aPKC coding sequence.

Received 8 January; accepted 12 February 2003; doi:10.1038/nature01486.

Published online 9 March 2003.

- Kuchinke, U., Grawe, F. & Knust, E. Control of spindle orientation in *Drosophila* by the Par-3-related PDZ-domain protein Bazooka. *Curr. Biol.* **8**, 1357–1365 (1998).
- Schober, M., Schaefer, M. & Knoblich, J. A. Bazooka recruits Inscutable to orient asymmetric cell divisions in *Drosophila* neuroblasts. *Nature* **402**, 548–551 (1999).
- Wodarz, A., Ramrath, A., Kuchinke, U. & Knust, E. Bazooka provides an apical cue for Inscutable localization in *Drosophila* neuroblasts. *Nature* **402**, 544–547 (1999).
- Petronczki, M. & Knoblich, J. A. DmPAR-6 directs epithelial polarity and asymmetric cell division of neuroblasts in *Drosophila*. *Nature Cell Biol.* **3**, 43–49 (2001).
- Wodarz, A., Ramrath, A., Grimm, A. & Knust, E. *Drosophila* atypical protein kinase C associates with Bazooka and controls polarity of epithelia and neuroblasts. *J. Cell Biol.* **150**, 1361–1374 (2000).
- Hirata, J., Nakagoshi, H., Nabeshima, Y. & Matsuzaki, F. Asymmetric segregation of the homeodomain protein Prospero during *Drosophila* development. *Nature* **377**, 627–630 (1995).
- Knoblich, J. A., Jan, L. Y. & Jan, Y. N. Asymmetric segregation of Numb and Prospero during cell division. *Nature* **377**, 624–627 (1995).
- Spana, E. P. & Doe, C. Q. The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. *Development* **121**, 3187–3195 (1995).
- Rhyu, M. S., Jan, L. Y. & Jan, Y. N. Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–491 (1994).
- Ishikawa-Kataoka, H., Skeath, J. B., Nabeshima, Y., Doe, C. Q. & Matsuzaki, F. Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. *Nature* **390**, 625–629 (1997).
- Shen, C. P., Jan, L. Y. & Jan, Y. N. Miranda is required for the asymmetric localization of Prospero during mitosis in *Drosophila*. *Cell* **90**, 449–458 (1997).
- Lu, B., Rothenberg, M., Jan, L. Y. & Jan, Y. N. Partner of Numb colocalizes with Numb during mitosis and directs Numb asymmetric localization in *Drosophila* neural and muscle progenitors. *Cell* **95**, 225–235 (1998).
- Ohno, S. Intercellular junctions and cellular polarity: the PAR-aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr. Opin. Cell Biol.* **13**, 641–648 (2001).
- Ohshiro, T., Yagami, T., Zhang, C. & Matsuzaki, F. Role of cortical tumour-suppressor proteins in asymmetric division of *Drosophila* neuroblast. *Nature* **408**, 593–596 (2000).
- Peng, C. Y., Manning, L., Albertson, R. & Doe, C. Q. The tumour-suppressor genes *lgl* and *dlg* regulate

- basal protein targeting in *Drosophila* neuroblasts. *Nature* **408**, 596–600 (2000).
16. Kalmes, A., Merdes, G., Neumann, B., Strand, D. & Mechler, B. M. A serine-kinase associated with the p127-(l2)gl tumour suppressor of *Drosophila* may regulate the binding of p127 to nonmuscle myosin II heavy chain and the attachment of p127 to the plasma membrane. *J. Cell Sci.* **109**, 1359–1368 (1996).
 17. Strand, D., Raska, I. & Mechler, B. M. The *Drosophila* lethal(2)giant larvae tumour suppressor protein is a component of the cytoskeleton. *J. Cell Biol.* **127**, 1345–1360 (1994).
 18. Musch, A. *et al.* Mammalian homolog of *Drosophila* tumour suppressor lethal (2) giant larvae interacts with basolateral exocytic machinery in Madin-Darby canine kidney cells. *Mol. Biol. Cell* **13**, 158–168 (2002).
 19. Kagami, M., Toh-e, A. & Matsui, Y. Sro7p, a *Saccharomyces cerevisiae* counterpart of the tumour suppressor l(2)gl protein, is related to myosins in function. *Genetics* **149**, 1717–1727 (1998).
 20. Lehman, K., Rossi, G., Adamo, J. E. & Brenwald, P. Yeast homologues of tomosyn and lethal giant larvae function in exocytosis and are associated with the plasma membrane SNARE Sec9. *J. Cell Biol.* **146**, 125–140 (1999).
 21. Strand, D. *et al.* The *Drosophila* lethal(2)giant larvae tumour suppressor protein forms homo-oligomers and is associated with nonmuscle myosin II heavy chain. *J. Cell Biol.* **127**, 1361–1373 (1994).
 22. Drier, E. A. *et al.* Memory enhancement and formation by atypical PKM activity in *Drosophila melanogaster*. *Nature Neurosci.* **5**, 316–324 (2002).
 23. Strand, D. *et al.* A human homologue of the *Drosophila* tumour suppressor gene l(2)gl maps to 17p11.2–12 and codes for a cytoskeletal protein that associates with nonmuscle myosin II heavy chain. *Oncogene* **11**, 291–301 (1995).
 24. Schaefer, M., Shevchenko, A. & Knoblich, J. A. A protein complex containing inscuteable and the galpha-binding protein pins orients asymmetric cell divisions in *Drosophila*. *Curr. Biol.* **10**, 353–362 (2000).
 25. Lin, D. *et al.* A mammalian PAR-3–PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. *Nature Cell Biol.* **2**, 540–547 (2000).
 26. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
 27. Schaefer, M., Petronczki, M., Dorner, D., Forte, M. & Knoblich, J. A. Heterotrimeric G proteins direct two modes of asymmetric cell division in the *Drosophila* nervous system. *Cell* **107**, 183–194 (2001).
 28. Schulze, K. L. & Bellen, H. J. *Drosophila* syntaxin is required for cell viability and may function in membrane formation and stabilization. *Genetics* **144**, 1713–1724 (1996).
 29. Kiehart, D. P. & Feghali, R. Cytoplasmic myosin from *Drosophila melanogaster*. *J. Cell Biol.* **103**, 1517–1525 (1986).
 30. Clemens, J. C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA* **97**, 6499–6503 (2000).

Acknowledgements We thank B. Dickson, M. Glotzer, A. Wutz and D. Berdnik for comments on the manuscript; Y. N. Jan, D. Kiehart, F. Matsuzaki, C. Thummel, the Developmental Studies Hybridoma Bank (DSHB) and the Bloomington *Drosophila* Stock Center for antibodies and fly stocks; and M. Petronczki for generating mouse anti-Par-6. Work in the laboratory of J.A.K. is supported by Boehringer Ingelheim and the Wiener Wirtschafts Foerderungsfond (WWFF).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.K. (e-mail: knoblich@nt.imp.univie.ac.at).

Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8

Helen Walden*, Michael S. Podgorski* & Brenda A. Schulman*†

* Department of Structural Biology, and † Department of Genetics/Tumor Cell Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

Post-translational modification by ubiquitin-like proteins (Ublps) is an essential cellular regulatory mechanism^{1–3}. The Ublp NEDD8 regulates cell division, signalling and embryogenesis^{4–6}. Ublps are conjugated to their targets by the sequential action of E1, E2 and often E3 enzymes³. Each Ublp has a dedicated E1, or activating enzyme, that initiates its conjugation cascade^{1,3,7–10}. First, E1 associates with the Ublp and catalyses adenylation of the carboxy terminus of the Ublp. Second, E1 forms a thioester between its catalytic cysteine and the Ublp. Next, E1 is loaded with a second Ublp molecule, adenylating the C terminus of this second Ublp while still carrying the first thioester-bound Ublp. Last, E1 binds E2 and promotes Ublp

transfer to the catalytic cysteine of E2. We report here the structure and mutational analysis of human APPBP1–UBA3, the heterodimeric E1 enzyme for NEDD8 (ref. 11). Each E1 activity is specified by a domain: an adenylation domain resembling bacterial adenylating enzymes¹², an E1-specific domain organized around the catalytic cysteine, and a domain involved in E2 recognition resembling ubiquitin. The domains are arranged around two clefts that coordinate protein and nucleotide binding so that each of E1's reactions drives the next, in an assembly-line fashion.

E1 activity^{7–10} (Fig. 1) is essential for Ublp conjugation¹³. The E1 for ubiquitin is a protein with a relative molecular mass of about 110,000 (M_r , 110K), and E1 enzymes for the Ublps NEDD8 and SUMO are heterodimers of about 110K. One protein (APPBP1 for NEDD8) is homologous to the amino-terminal half of the E1 for ubiquitin, and the other protein (human Uba3, homologue of yeast Uba3 and hereafter referred to as UBA3, for NEDD8) is homologous to the C-terminal half. The C-terminal half contains both the nucleotide-binding motif involved in adenylation and the catalytic cysteine involved in the thioester intermediate and Ublp transfer to E2 (refs 11, 14–16) (Fig. 2; see also Supplementary Fig. 1). Both halves of E1 share a region of sequence homology also found in the bacterial enzymes MoeB and ThiF¹⁵ (Supplementary Fig. 1). MoeB and ThiF adenylate the C termini of MoaD and ThiS, components of metabolic enzymes that are structural homologues of Ublps^{1,12}. The previous MoeB–MoaD adenylate structure provided insights into the E1 adenylation reaction¹²; however, the structural bases for E1 selection of Ublps, E1–Ublp thioester formation, and E1-mediated transfer of Ublp to E2 remain unknown. To understand these functions, we determined the 2.6 Å APPBP1–UBA3 crystal structure (Supplementary Table 1 and Supplementary Fig. 2). Here we describe the structure in order of the three E1 activities, adenylation, thioester bond formation and Ublp transfer to E2.

The overall APPBP1–UBA3 structure contains a $\sim 50 \text{ Å} \times 30 \text{ Å} \times 20 \text{ Å}$ groove, with the adenylation domain as the base and the catalytic cysteine-containing domain and ubiquitin-like domain as the walls (Fig. 2). The groove is divided into two clefts by the L7 loop of UBA3, leading from the adenylation domain to the catalytic cysteine. Each cleft is sufficiently large to accommodate either two molecules of NEDD8 or one molecule of E2. Cleft 1 has the nucleotide-binding pocket as its base, whereas the walls are formed by the UBA3 portion of the catalytic cysteine domain, and by the ubiquitin-like domain. Cleft 2 has UBA3 β -strands g–j as its base, and the catalytic cysteine domain portion of APPBP1 as its only wall. The catalytic cysteine of UBA3 faces cleft 2 (Fig. 2).

The adenylation domain comprises APPBP1 residues 6–168 and 486–534, and UBA3 residues 12–210 and 290–347. In this domain, APPBP1 and UBA3 are structurally homologous to each other and to MoeB¹², including regions of low sequence homology (1.6 Å root-mean-square deviation (r.m.s.d.) over 412 residues). The structures

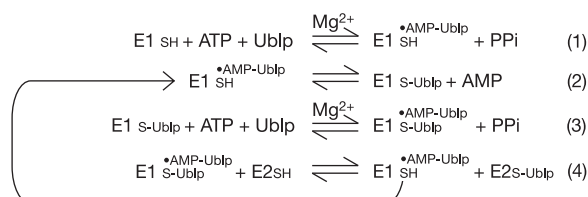


Figure 1 E1 reaction scheme^{1,3,7–10}, with noncovalent complexes indicated with a dot, and covalent complexes indicated with a hyphen. SH refers to the free form of the catalytic cysteine, S-Ublp refers to the thioester between the catalytic cysteine and the Ublp C terminus, and PPi refers to inorganic pyrophosphate.

adopt a variant Rossmann fold, with eight α -helices surrounding a mixed eight-stranded β -sheet (Fig. 2; see also Supplementary Fig. 1). The adenylation domain interface between APPBP1 and UBA3 resembles that in the crystallographic MoeB homodimer, burying 5,700 Å². A total of 63 of the 147 interface residues are identical in other E1 enzymes (Fig. 2; see also Supplementary Fig. 1). Unlike the MoeB homodimer that has two adenylation active sites, APPBP1–UBA3 has only one, centred around the canonical nucleotide-binding motif of UBA3. Two features of APPBP1, conserved in other E1 enzymes, preclude an adenylation reaction. First, APPBP1 lacks a nucleotide-binding motif. Second, a 4-helix bundle (helices 21–24) packs against the Rossmann fold, further prohibiting nucleotide binding.

To identify determinants of adenylation, we modelled NEDD8 binding (Fig. 3a). After superposition of the APPBP1–UBA3 adenylation domain with MoeB–MoaD adenylate¹², the structure of NEDD8 (ref. 17) was modelled into the site of MoaD. Both clefts are involved in NEDD8 adenylate binding, with nucleotide binding the floor of cleft 1, and NEDD8 contacting the floor of cleft 2 (Fig. 3a). Residues involved in nucleotide binding and catalysis of the adenylation reaction, revealed by MoeB–MoaD adenylate¹², are conserved in the APPBP1–UBA3 structure. UBA3 residues Leu 145, Ile 127, Gly 55 and Ile 54 contact the adenine, Asp 79 contacts the ribose ring, and Asp 146 coordinates Mg²⁺. The Arg 15 residue of APPBP1 contacts ATP's γ -phosphate.

The model predicts that exposed hydrophobic residues from UBA3 β -strands g–j and 3₁₀-helix E contact NEDD8. A hydrophobic patch (Tyr 321, Tyr 331, Tyr 333 and Phe 335), conserved as hydrophobic residues in E1 of ubiquitin, contacts the Leu 8, Ile 44, Val 70 hydrophobic patch of NEDD8, which is identical in ubiquitin (Fig. 3a). These residues in ubiquitin are essential for yeast viability¹⁸, and mutation of Leu 8, or Leu 8 plus Val 70, reduces ubiquitin-conjugate formation¹⁹. Arg 42, important for ubiquitin-adenylate formation²⁰ and conserved in NEDD8, forms a hydrogen bond with UBA3 Tyr 321. We tested our model by investigating the effects of mutations on NEDD8 adenylation. Mutation of the UBA3 residues Leu 145 and Asp 146 in the nucleotide-binding site, and Tyr 331, Tyr 333 and Phe 335 in the putative NEDD8-binding site, impedes NEDD8 adenylation, agreeing with the model (Supplementary Fig. 3a). Mutation of Ile 127, which contributes only 3% of predicted contacts with AMP, has a lesser effect. Mutation of residues predicted not to contact nucleotide or NEDD8 has no effect.

The model suggests a rationale for part of the preference of E1 enzymes for their cognate Ubblps. The only known determinant of E1 specificity in Ubblps is residue 72, which is alanine in NEDD8 and arginine in ubiquitin^{17,20}. The corresponding residue in SUMO family members is glutamate or a glutamine. In the model, the Ala 72 of NEDD8 interacts with the hydrophobic Leu 206 and Tyr 207 residues of UBA3 (Fig. 3a). The nature of the side chains

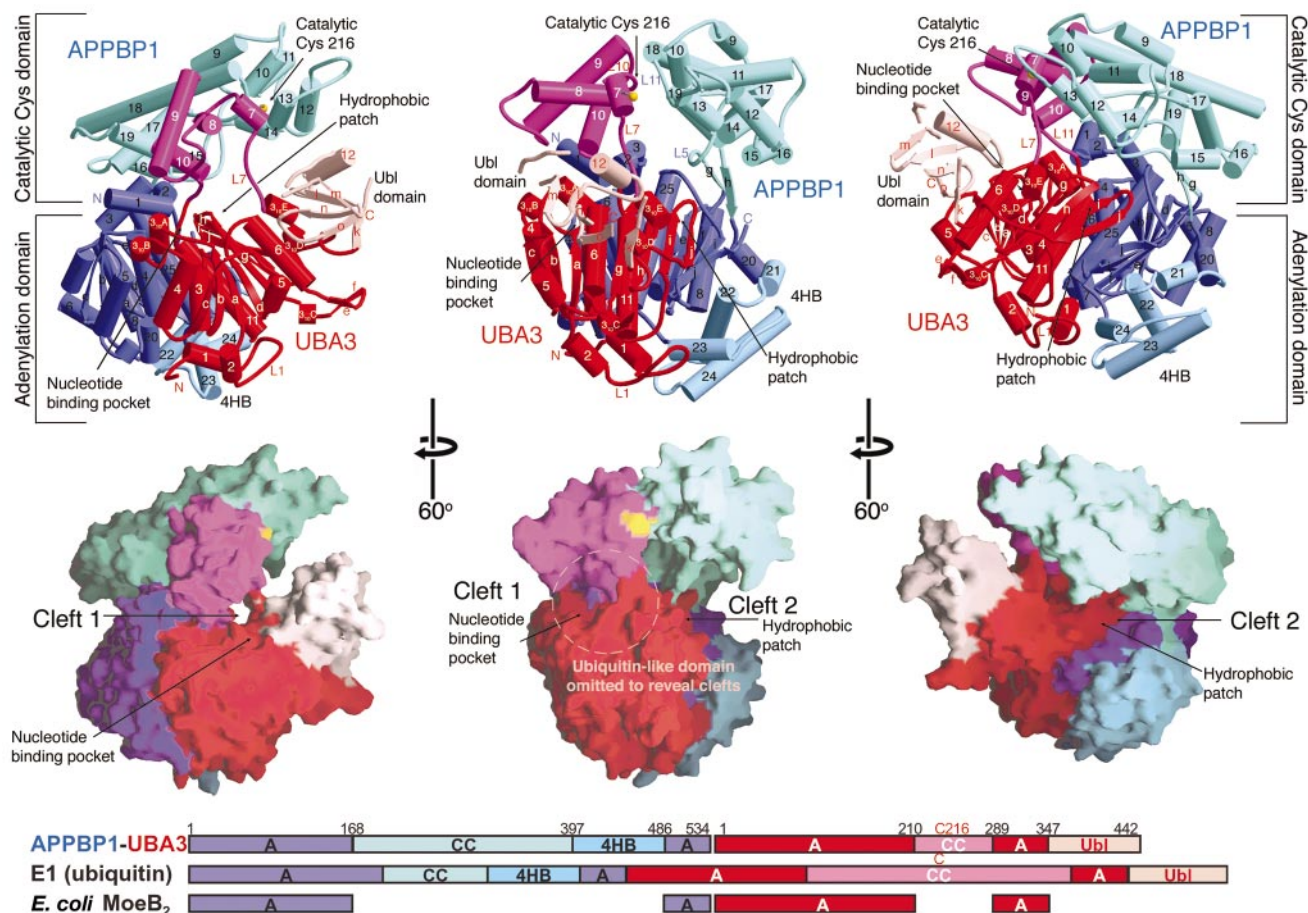


Figure 2 Structure of the APPBP1–UBA3 complex. Three views of the structure, separated by 60° clockwise rotations in the y axis, with structural elements shown in the top panels, and surface representations below. APPBP1 is blue, whereas UBA3 is red, with domains shaded differently. Catalytic Cys 216 is yellow. The primary sequences of APPBP1–UBA3, the E1 of ubiquitin and the crystallographic MoeB dimer¹² are shaded

according to the structure in the schematic diagram (bottom). A, adenylation domain; CC, catalytic cysteine domain; 4HB, 4-helix bundle in APPBP1; Ubl, ubiquitin-like domain in UBA3. The adenylation, catalytic cysteine and ubiquitin-like domains together contribute to a broad, deep groove in the structure that is divided into two clefts, which coordinate nucleotide and protein binding.

corresponding to Leu 206 and Tyr 207 in other E1 enzymes correlates with the nature of the side chain at position 72 in other Ubblps. Leu 206 and Tyr 207 are replaced by an aspartate in the E1 of ubiquitin, and by a lysine in the E1 of SUMO molecules. We speculate that the Arg 72 of ubiquitin forms a salt bridge with the aspartate of its E1, and that SUMO's glutamate or glutamine forms a salt bridge or hydrogen bond with the lysine of its E1. Mutation of Leu 206 and Tyr 207 of UBA3 to aspartate, corresponding to ubiquitin's E1 sequence, greatly diminishes NEDD8 adenylation (Supplementary Fig. 3a). This mutation is not sufficient for APPBP1–UBA3 to promote adenylation of ubiquitin, however, suggesting the existence of other determinants of specificity among E1 enzymes.

A second domain, entirely helical, comprises E1-specific insertions in the sequences of the adenylation domain portions of both APPBP1 and UBA3 (Fig. 2; see also Supplementary Fig. 1). The focus

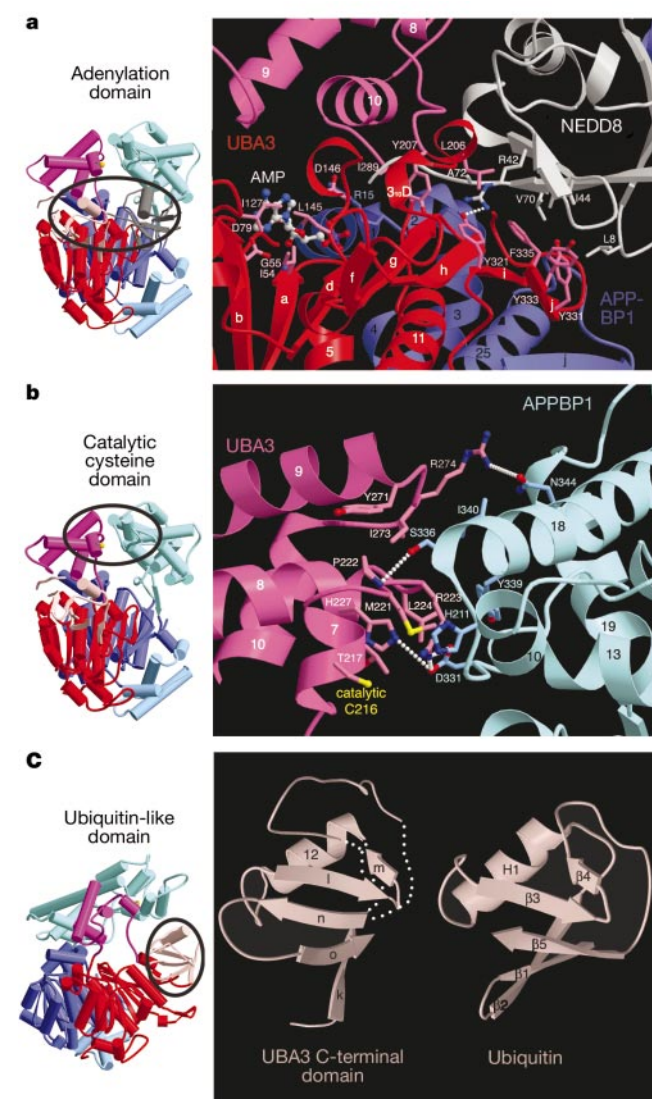


Figure 3 Close-up views of the adenylation, catalytic cysteine and ubiquitin-like domains. **a**, Model of APPBP1–UBA3 (blue and red, respectively) complexed with AMP and NEDD8 (grey). After superposition of the APPBP1–UBA3 adenylation domain with MoeB–MoaD adenylate¹², the structure of NEDD8 (ref. 17) was modelled into the site occupied by MoaD. Nucleotide and NEDD8 interacting residues are labelled. **b**, Close-up view of UBA3's catalytic Cys 216 nearby residues involved in intermolecular interactions. **c**, The C-terminal domain of UBA3 adopts the ubiquitin fold.

of this domain is the catalytic Cys 216 of UBA3, which forms a thioester with NEDD8 and transfers NEDD8 to its E2 (refs 11, 16). Although the insertions in family members paralleling APPBP1 are shorter and those in family members paralleling UBA3 are correspondingly longer, they are predicted to be entirely helical, so the overall structure of this domain will probably be similar in other E1 enzymes. The corresponding region of MoeB is an eight-residue mobile loop including a cysteine not involved in catalysis¹².

The elongated catalytic cysteine domain contains 11 helices from APPBP1 and 4 from UBA3 that radiate towards or away from Cys 216. The catalytic cysteine domain interface buries 900 Å² and involves UBA3 helices 8 and 10 and loop L10 packing against APPBP1 helix 18 and loops L6 and L11 (Fig. 3b). Two hydrophobic clusters stabilize this interface: (1) APPBP1 Ile 340 and Tyr 339 interact with UBA3 Arg 223, Leu 224, Tyr 271 and Ile 273; (2) APPBP1 His 211 and Ile 333 pack against UBA3 Met 221. This interface also contains several hydrogen bonds and salt bridges: between APPBP1 Asn 344 and UBA3 Arg 274, between APPBP1 Ser 336 and the backbone oxygen of UBA3 Pro 222, and between APPBP1 Asp 331 and UBA3 His 227 and Arg 223. Conservation of side chains suggests that this interface will be similar in other E1 enzymes (Supplementary Fig. 1). Mutation of APPBP1 Asp 331 to alanine reduces the ability of APPBP1–UBA3 to form a thioester with NEDD8 (Supplementary Fig. 3b), confirming this interface's important role.

The structure raises the question of the mechanism of E1–Ubblp thioester bond formation, which would require that Cys 216 and the C terminus of NEDD8 be in close proximity. The Cys 216 thiol is 28 Å away from the C terminus of NEDD8 in our model. This suggests that juxtaposition of Cys 216 and the C terminus of NEDD8 would require a conformational change. As the C-terminal residues of NEDD8 and other Ubblps are flexible¹⁷, NEDD8 could

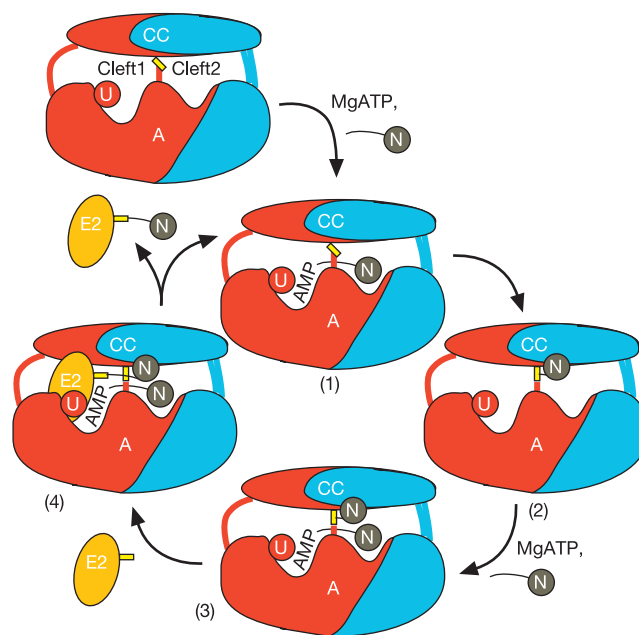


Figure 4 Structure-based model for E1 reactions. APPBP1 is cyan, UBA3 is red, NEDD8 is grey and UBC12 is orange. The catalytic cysteine is a yellow bar. A, APPBP1–UBA3 adenylation domain; CC, catalytic cysteine domain; U, ubiquitin-like domain. (1) The adenylation reaction involves nucleotide binding to cleft 1 and NEDD8 binding to cleft 2; (2) the catalytic cysteine, facing NEDD8, forms a thioester with NEDD8; (3) a second molecule of NEDD8 is adenylated; (4) occupation of the nucleotide-binding pocket facilitates binding of E2 in cleft 1, promoting transfer of the E1 thioester-linked Ubblp to E2.

remain bound to the hydrophobic patch as in our model, and the gap would be reduced to only 10 Å if the C terminus of NEDD8 extended directly towards Cys 216. The remainder of the gap might be closed by a change in the relative orientations of the adenylation and catalytic cysteine domains, or by a local conformation change. Formation of the E1-Ublp thioester probably requires deprotonation of the catalytic cysteine of E1, and formation of the E2-Ublp thioester probably requires deprotonation of E2's catalytic cysteine. Histidines are often general-base catalysts. APPBP1 His 211 and UBA3 His 227 are 6.8 and 9.8 Å away from Cys 216, respectively. Not only are these histidines distal from the catalytic cysteine, but they also participate in the APPBP1-UBA3 interface. The UBA3 Cys 216 side chain is exposed in the active site except for Thr 217, 3.75 Å away (Fig. 3b). A possible role for Thr 217 is suggested from its conservation in E1 enzymes for NEDD8, ubiquitin and SUMO. The corresponding residue is valine in MoeB, whose cysteine is non-catalytic. Thr 217 mutation to alanine decreases thioester formation between APPBP1-UBA3 and NEDD8 (Supplementary Fig. 3b), without affecting NEDD8 adenylate formation (Supplementary Fig. 3a).

Unexpectedly, the 95-residue C-terminal domain of UBA3 superimposes with the structure of ubiquitin (r.m.s.d. of 1.9 Å over 50 residues), except for β -strand α , which is in the opposite direction to the equivalent strand in ubiquitin (Fig. 3c). Notably, many proteins in the ubiquitin pathway contain ubiquitin-like domains². We hypothesized that the C-terminal domain might be involved in E2 binding for three reasons. First, the C-terminal domain borders cleft 1, which contains the nucleotide-binding site. Transfer of E1's thioester-bound Ublp to E2 is stimulated by E1 binding to Ublp adenylate²¹. Second, ubiquitin-like domains are found in some E3 enzymes, and E3s are known to bind E2s^{22,23}. Third, the Ublp Smt3p competes with E1 for binding to the E2 Ubc9p (ref. 24). Indeed, a mutant APPBP1-UBA3 lacking the C-terminal domain is defective at associating with UBC12 and transferring NEDD8 to UBC12 (Supplementary Fig. 3c, d). Mutation of UBA3 Ile 289, exposed in cleft 1, also diminishes NEDD8 transfer to UBC12 (Supplementary Fig. 3c).

The APPBP1-UBA3 structure provides a framework for understanding E1's multiple biochemical activities^{3,7-10,21} (Fig. 4). Although different domains carry out adenylation, thioester bond formation and E2 association, together they form two clefts that coordinate nucleotide and protein binding so that each one of E1's reactions drives the next. Both clefts are involved in NEDD8 adenylation. The subsequent thioester bond formation is probably facilitated by the catalytic cysteine facing cleft 2, occupied by NEDD8. The next step in the cascade involves E1 binding to E2, which is enhanced by nucleotide binding to E1 (refs 3, 10, 21). Our results suggest that both the E2 and nucleotide-binding sites are in cleft 1. The bound nucleotide may cause structural rearrangements affecting E2 binding, or it may occupy part of the E2-binding site. The structure suggests that E1 enzymes drive the initial steps of Ublp transfer cascades in an assembly-line fashion. □

Methods

Cloning, expression and purification

APPBP1, UBA3, UBC12 and NEDD8 were purchased as IMAGE clones from Research Genetics/Incyte Genomics (identification numbers 2964638, 2371074, 3344837 and 4801390, respectively). All experiments described here used a truncated version of NEDD8 terminating at Gly 76, representing the active form. Mutations were generated by polymerase chain reaction (PCR), and the entire coding sequence was verified by automated sequencing. Glutathione S-transferase (GST)-APPBP1 and UBA3 were expressed together in BL21(DE3) (Novagen) from a bicistronic vector based on pGEX4T1 (Pharmacia). GST-UBC12 was expressed similarly from pGEX4T1, and GST-NEDD8 and GST-kNEDD8, which has an N-terminal protein kinase A (PKA) phosphorylation site, were expressed in BL21(DE3)/RIL (Novagen) from pGEX4T1 and pGEX2TK. All were purified by glutathione affinity chromatography. The GST fusions of APPBP1-UBA3 used for crystallization, and of UBC12, NEDD8 and kNEDD8 were cleaved by thrombin. APPBP1-UBA3 and NEDD8 were further purified by anion exchange and gel filtration chromatography. UBC12 and tNEDD8 were further purified by cation-exchange

chromatography. The APPBP1-UBA3 used for crystallization was concentrated to 10–15 mg ml⁻¹, and proteins for biochemical assays were concentrated to 0.5–5 mg ml⁻¹. The C-terminal domain deletion mutant terminates after residue Gln 348 of UBA3.

Crystallization and structure determination

Crystals of native and selenomethionine-labelled APPBP1-UBA3 grew in 15–17% PME5500, 0.1 M HEPES, 10 mM dithiothreitol (DTT), pH 7.0 at 4 °C after multiple rounds of microseeding and macroseeding. They form with two complexes in the asymmetric unit in the space group P2₁2₁2₁ with $a = 92.4$ Å, $b = 123.6$ Å, $c = 198.2$ Å. Multiwavelength anomalous dispersion (MAD) data were collected at National Synchrotron Light Source (NSLS) beamline X25 on crystals flash-frozen in liquid nitrogen in crystallization buffer supplemented with an additional 2% PME5500 and 25% MPD. Data were processed using DENZO/SCALEPACK²⁵. A total of 18 out of the expected 34 Se sites were identified by SOLVE²⁶. Twenty-four sites, including an additional six non-crystallographic symmetry (NCS)-related sites, were improved by SHARP²⁷, and the phases were improved by solvent flattening using DM and 2-fold NCS-averaging to give a mean figure-of-merit of 0.8 to 2.8 Å. An $F_o - F_c$ map revealed density at 7 σ consistent with one tetrahedrally coordinated zinc in each heterodimer. Representative portions of the experimental electron density map are shown in Supplementary Fig. 2. The model was built manually in the program O²⁸ and refined with CNS²⁹ using isomorphous native data to 2.6 Å, collected at Advanced Light Source (ALS) beamline 5.0.2, with 5% removed to calculate R_{free} . The model contains APPBP1 residues 6–534 and UBA3 residues 12–354, 361–384 and 426–441, with 15 additional residues built as polyalanine owing to poor electron density in this region. NEDD8 and AMP were modelled onto the APPBP1-UBA3 structure in O, first by least-squares superposition of MoeB-MoaD adenylate¹² with UBA3, followed by least-squares superposition of NEDD8¹⁷ with MoaD. The fit was improved by energy minimization in CNS.

Biochemical assays

NEDD8 adenylate formation was monitored by following attachment of ³²P-labelled AMP to NEDD8. A total of 40 nM GST-APPBP1-UBA3 or mutant complex, 12 μ M NEDD8 and 10 μ M [α -³²P]ATP was incubated for 5 min in 6 μ l of 50 mM Tris, 50 mM NaCl, 10 mM MgCl₂, pH 7.6. Reactions were terminated by the addition of SDS-polyacrylamide gel electrophoresis (PAGE) sample buffer and the ³²P-labelled NEDD8 adenylate was resolved from free ³²P-labelled ATP by SDS-PAGE. UBA3-NEDD8 thioester bond formation was monitored by following formation of the reducible complex of UBA3 with kNEDD8 phosphorylated at an N-terminal PKA site with [γ -³²P]ATP³⁰. A total of 60 nM GST-APPBP1-UBA3 or mutant complex and 0.3 μ M ³²P-labelled kNEDD8 was incubated for 2 min in 6 μ l of 50 mM Tris, 50 mM NaCl, 10 mM MgCl₂, 10 μ M ATP, pH 7.6. Reactions were terminated by the addition of non-reducing SDS-PAGE sample buffer, and the ³²P-labelled UBA3-kNEDD8 was resolved from free ³²P-labelled kNEDD8 by non-reducing SDS-PAGE. UBC12-NEDD8 thioester bond formation was monitored similarly, with 10 μ M UBC12 included in the reaction mix. APPBP1-UBA3 binding to UBC12 was monitored by a native gel mobility shift assay²⁴. A total of 2.5 μ M APPBP1-UBA3 or mutant complex was incubated in the presence of increasing amounts of UBC12 (0.312, 0.625, 1.25, 2.5 and 5 μ M) in 10 μ l volumes in 25 mM Tris, 150 mM NaCl, 5% glycerol, 5 mM DTT, pH 7.6 for 1 h. Free and UBC12-bound APPBP1-UBA3 were separated on a 4.5% polyacrylamide gel (acrylamide/bis; 37.5:1) in 90 mM Tris-borate, 2% glycerol, pH 8.0, and visualized with Coomassie staining.

Received 10 October 2002; accepted 21 January 2003; doi:10.1038/nature01456.

- Hochstrasser, M. Biochemistry. All in the ubiquitin family. *Science* **289**, 563–564 (2000).
- Jentsch, S. & Pyrowolakis, G. Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol.* **10**, 335–342 (2000).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Lammer, D. et al. Modification of yeast Cdc53p by the ubiquitin-related protein Rub1p affects function of the SCFCdc4 complex. *Genes Dev.* **12**, 914–926 (1998).
- Pozo, J. C., Timpte, C., Tan, S., Callis, J. & Estelle, M. The ubiquitin-related protein RUB1 and auxin response in *Arabidopsis*. *Science* **280**, 1760–1763 (1998).
- Kurz, T. et al. Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **295**, 1294–1298 (2002).
- Haas, A. L., Warm, J. V., Herskko, A. & Rose, I. A. Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *J. Biol. Chem.* **257**, 2543–2548 (1982).
- Haas, A. L. & Rose, I. A. The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *J. Biol. Chem.* **257**, 10329–10337 (1982).
- Herskko, A., Heller, H., Elias, S. & Ciechanover, A. Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* **258**, 8206–8214 (1983).
- Haas, A. L., Bright, P. M. & Jackson, V. E. Functional diversity among putative E2 isozymes in the mechanism of ubiquitin-histone ligation. *J. Biol. Chem.* **263**, 13268–13275 (1988).
- Osaka, F. et al. A new NEDD8-ligating system for cullin-4A. *Genes Dev.* **12**, 2263–2268 (1998).
- Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature* **414**, 325–329 (2001).
- Finley, D., Ciechanover, A. & Varshavsky, A. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. *Cell* **37**, 43–55 (1984).
- Hatfield, P. M., Callis, J. & Vierstra, R. D. Cloning of ubiquitin activating enzyme from wheat and expression of a functional protein in *Escherichia coli*. *J. Biol. Chem.* **265**, 15813–15817 (1990).
- Johnson, E. S., Schwenhorst, I., Dohmen, R. J. & Blobel, G. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aosl1p/Uba2p heterodimer. *EMBO J.* **16**, 5509–5519 (1997).
- Liakopoulos, D., Doenges, G., Matuschewski, K. & Jentsch, S. A novel protein modification pathway related to the ubiquitin system. *EMBO J.* **17**, 2208–2214 (1998).
- Whitby, F. G., Xia, G., Pickart, C. M. & Hill, C. P. Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. *J. Biol. Chem.* **273**, 34983–34991 (1998).

18. Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M. & Hicke, L. Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* **276**, 30483–30489 (2001).
19. Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M. & Pickart, C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl Acad. Sci. USA* **93**, 861–866 (1996).
20. Burch, T. J. & Haas, A. L. Site-directed mutagenesis of ubiquitin. Differential roles for arginine in the interaction with ubiquitin-activating enzyme. *Biochemistry* **33**, 7300–7308 (1994).
21. Pickart, C. M., Kasperk, E. M., Beal, R. & Kim, A. Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). *J. Biol. Chem.* **269**, 7115–7123 (1994).
22. Kitada, T. *et al.* Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
23. Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumour suppressor function. *Science* **284**, 455–461 (1999).
24. Bencsath, K. P., Podgorski, M. S., Pagala, V. R., Slaughter, C. A. & Schulman, B. A. Identification of a multifunctional binding site on Ubc9p required for Smt3p conjugation. *J. Biol. Chem.* **277**, 47938–47945 (2002).
25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **176**, 307–326 (1997).
26. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
27. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
30. Scheffner, M., Huijbregtse, J. M., Vierstra, R. D. & Howley, P. M. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495–505 (1993).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

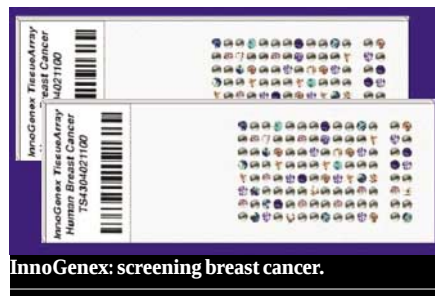
Acknowledgements We thank P. Jeffrey, D. Minor, J. Berger, J. Huijbregtse and C. Ross for advice and assistance; M. Becker and staff at National Synchrotron Light Source Beamline X25, and T. Earnest, G. McDermott and staff at Advanced Light Source Beamline 5.0.2 for help with data collection; and P. Murray, T. Izard and L. Hendershot for comments on the manuscript. This work was supported by ALSAC, the NCI Cancer Center (CORE), and a Pew Scholar in Biomedical Sciences Award to B.A.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.A.S.
(e-mail: Brenda.schulman@stjude.org). Coordinates have been deposited in the Protein Data Bank under accession code 1NGV.

Screen breaks

Focusing on high throughput.



InnoGenex: screening breast cancer.

Human breast cancer TMAs

InnoGenex www.innogenex.com

Diagnosis and prognosis

These tissue microarrays (TMAs) for human breast cancer can be used in high-throughput screening of *in situ* gene expression for target validation and differential gene expression profiling. A variety of tumours and different grades of tumours are represented on a single slide. Paired normal and abnormal samples from the same patient are also represented in the TMAs. Hence, gene expression can be correlated with normal breast tissue and breast cancer tissue, or with different stages of cancer development. The tissues are fixed in neutral buffered formalin, embedded in paraffin and sectioned at approximately 3–5 μm section thickness. They are validated for use in immunohistochemistry and may also be used for *in situ* hybridization assays. Low-density TMAs (20–50 elements) and high-density TMAs (100–200 elements) are available. The slides are provided with a detailed map showing patient information and type of breast cancer.

CellProbe

Beckman Coulter www.beckmancoulter.com

Streamlined assay for apoptosis detection

This is the first in Beckman Coulter's new line of high-throughput fluorogenic kits for cell-based assays. CellProbe is a caspase 3.7 assay kit designed for apoptosis detection. CellProbe can be automated on Biomek liquid-handling workstations, making its use feasible in high-throughput drug discovery applications such as primary and secondary screening. The plate-based assay kit requires no protease inhibitors or mixing. By eliminating cell lysis and washing steps, liquid handling is streamlined for easy automation. High specificity enables researchers to easily measure activity in whole cells for detection of small changes in enzyme activity. The kit is available with protocols for 96- and 384-well assay formats.

Colorware

Kimble-Kontes www.kimble-kontes.com

Brighten your benchtop

Kimble-Kontes offers a range of easily identifiable colour-coded glassware that will reduce 'roaming' from lab to lab and minimize the possibility of cross-contamination caused by glassware being mixed between laboratories. The range includes low-form beakers, Erlenmeyer flasks and volumetric flasks printed with easy-to-read, durable colour ceramic enamel graduations and matte marking spots in a choice of 'sunny yellow', 'raging red', 'bright blue' or 'cool green'.

Application notes

Advanced ChemTech

www.advancedchemtech.com

For online bibliophiles

ChemTech's online bibliography contains applications and technical information on both solution and solid-phase synthesis methods in combinatorial, organic and peptide chemistry. The application notes can be downloaded from the company's website as pdf documents. Topics covered include solution phase synthesis with scavenger resins, removal of allyl and alloc protecting groups, a comparison of Nsc versus Fmoc as a base-labile N^α -amine protecting group, solid-phase preparation of 2-piperazinones, and useful solvents for high-temperature solid-phase chemistry. Each application note provides a short introduction to the subject matter as well as a detailed experimental procedure, including equipment and materials required for the reaction described. The bibliography can be found at www.advancedchemtech.com/support/app_notes.htm

New drivers for EDAS 290

Kodak Scientific Imaging Systems

www.kodak.com

Gel documentation for Windows XP

Windows XP drivers are now available for Kodak's EDAS 290 2.1 mega-pixel digital CCD imaging system for electrophoresis gel documentation and routine point-and-shoot photography. The EDAS 290 LE (limited edition) and 290 FV (full version) allows users to preview and capture high-resolution images of their samples. The full version also includes Kodak's 1D image analysis software for ready calculation of molecular weight, mass, optical density and intensity measurements.



Microassays in maxi quantities: the OmniGrid 300.

OmniGrid 300

GeneMachines www.genemachines.com
Production-level microarrayer

The OmniGrid 300 microarrayer is designed to meet the demands of ultra-high-throughput production facilities. The instrument boasts a capacity of 308 slides and seventy-three 384-well plates. It also features automatic plate loading, flow-through sonication and an optional second wash/dry station to support a broader range of cleaning options.

Cellroll cultivation system

Integra Biosciences
www.integra-biosciences.com
Click and roll

This system automates the task of rolling cell cultivation bottles. It comprises a moisture-resistant bottle deck and removable control unit, and can be expanded to cope with up to 32 roller bottles. Cell cultivation can take place in both standard and CO₂ incubators. Additional decks may be added to the basic model in a vertical or horizontal formation using a simple click-in and roll mechanism. The Cellroll control unit enables the entire cell cultivation process to be pre-programmed through a non-volatile memory facility. The system can be fitted with a slow or rapid rotating motor. Velocity, angle of rotation, intervals between work cycles and length of cycles can be programmed to be automatically adjusted, allowing vibration-free and secure handling of sensitive cell lines. Applications include cultivation of adherent cells (such as vaccine production) and cells grown in suspension, such as insect cells.

GPCR screening

PerkinElmer www.perkinelmer.com
Assays to enhance GPCR screening

G protein coupled receptor (GPCR) screening is a primary focus of drug development. The Delfia kit is a non-radiometric GTP

binding assay that screens new agonists and confirms secondary assays in drug discovery by monitoring therapeutic-induced changes in GPCR activity. PerkinElmer also offers two luminescence assay systems, Britelite and Steadylite, for reporter gene assays used in functional GPCR screening applications. Britelite provides an ultra-bright signal to maximize sensitivity in high-throughput screening, while Steadylite features a long-lived signal, making it suitable for batch processing. Both offer one-plate preparation from cell culture to reagent addition and analysis, and are available in 96-, 384- or 1,536-well formats. Finally, Image Flash-Plate is a 384-shallow-well microplate coated with a scintillant that emits in the red region, combining ultra-miniaturization with high sensitivity.



Nunc's new microplate lids.

LowBot Lids

Nunc A/S www.nuncbrand.com
Keep a low profile

Nunc's LowBot low-profile microplate lids feature cut-away sides, designed to optimize robotic handling and barcode reading for standard microplates. Made of clear polystyrene, these lids are particularly suitable for 384-well, 1,536-well, and other low-profile design plates. External dimensions are 128 × 86 mm, and they come in sterile packs of 25 or 100.

Glass fibre and MacroFilter plates

Porvair Sciences www.porvair-sciences.com

Filtration microplates for all applications

Porvair's range of 48- and 96-well microplates incorporates a choice of glass fibre, nylon, nitrocellulose, PVDF or polyethylene supports, in well volumes ranging from 350 µl to 5 ml per well. Glass fibre filtration plates can be used in a number of high-throughput microfiltration applications, including cell harvesting, DNA separations, binding studies, plasmid isolation, general filtration and sample clean-up, and can filter 96 samples of up to 350 µl volume. MacroFilter plates feature medical grade polymer filters and can be

used to filter 48 samples of up to 5 ml. The polymer filters optimize sample integrity by minimizing leachates with a wide range of solvents. Each well on the MacroFilter plates has an individual drainage spout for 100% sample transfer with no crossover contamination.

MagneHis

Promega www.promega.com

Versatile choice for His-tagged proteins

The MagneHis protein purification system can be used to purify polyhistidine-tagged expressed proteins either manually or automatically in both small- and large-scale applications (1 ml to over 1 litre of culture). The system is particularly suitable for purification of recombinant proteins containing a His-tag at either the C or N terminus. MagneHis nickel particles are compatible with a variety of buffers. His-tagged proteins can be isolated directly from a crude cell lysate and no vacuum or centrifugation steps are required once cells have been lysed. The system has a binding capacity of 1 mg His-tagged protein per 1 ml of MagneHis nickel particles. Applications include protein-protein interactions, enzyme kinetics/activity or structural analysis.

Genesis Freedom

Tecan www.tecan.com

Armed for action

This modular system from Tecan is based on a liquid handling instrument that can be upgraded to a fully automated robotic workstation. It supports up to three different types of robotic arm on one system — a liquid handling arm, a robotic manipulator arm, and a pick-and-place arm — that provide up to nine different arm configurations. Access to space below the deck through openings on the work table allows integration of a range of other devices, including centrifuges and plate systems, all within easy reach of the system's robotic arms. The Genesis Freedom can be used in genomics (PCR set-up, genotyping), proteomics (MALDI spotting, sample preparation for flow cytometry), and drug discovery (ADME, high-throughput screening and toxicology studies).

HPF Millex syringe filters

Millipore www.millipore.com

For a clear solution

These syringe filters use two different media, a graduated glass fibre prefilter (10.0 to 0.7 µm) to remove larger particles, and a 0.45 µm membrane filter for fine filtration — to optimize throughput. Users can choose from the HPF Millex-HN filter, which is recommended for filtering solvents and has a nylon membrane with broad chemical compatibility, or the HPF Millex-HV filter, which has a Dura-pore low-protein binding membrane recommended for filtering proteinaceous solutions.



HPF Millex: for proteinaceous filtrations.

The filters are available in packs of 50 and 1,000 for individual sample filtration using a syringe, and in packs of 200 (8 × 25 tubes) for automated sample preparation with Zymark workstations. All filters feature a domed, rigid housing to prevent carousel jams.

Plant genome DNA

Promega/CropDesign www.promega.com

Genomic plant DNA free of PCR inhibitors

The Wizard Magnetic 96 DNA Plant System is an ultra-high-throughput plant genomic DNA purification based on protocols for plant DNA extraction developed by CropDesign NV of Gent, Belgium. The methodology was developed using Promega technology and the Tecan robotic platform. The protocol is capable of processing 24 plates in a 24-hour period on a Tecan Genesis robotic system, using proven MagneSil paramagnetic silica technology. This combination is easily amenable to automation allowing plant molecular biologists to save time and streamline their sample preparation processes.

NuGenesis SDMS Version 6

NuGenesis Technologies Corporation

www.nugenesis.com

Next-generation data management

Version 6 of the NuGenesis scientific data management system (SDMS) includes enhancements designed to make the system applicable through the entire life cycle of scientific data, and to facilitate easy integration within the existing technology structure. With version 6 NuGenesis aims to meet the needs of the scientists, IT, regulatory and business management.

These notes are compiled in the Nature office from information provided by the manufacturers. Material intended for publication can be sent to newproducts@nature.com.

Hitting the target

The route to new therapeutics often ends in costly failure. The secret of success is the rapid and accurate identification of drug targets with true potential, says Caitlin Smith.

For the pharmaceutical industry, the Human Genome Project has proved to be both a blessing and a curse.

Where potential drug targets were once hard to come by, the industry is now awash with them. This has left researchers with the unenviable challenge of sifting through the data in search of the elusive proteins that are instrumental in human disease.

Akin to seeking a needle in a haystack, this herculean task has boosted the importance of rapid screening technologies. Of the roughly 35,000 genes in the human genome, only a few have known functions. So the task of identifying and verifying a positive lead is key to effective drug development.

Drugs fail in the clinic for two basic reasons: they either don't work or they prove to be unsafe. "Both of these are often the direct result of sloppy early target validation," says David Szymkowski, director of biotherapeutics at the biopharmaceutical company Xencor in Monrovia, California.

Validation is a crucial step in the drug-discovery process. Most drugs are inhibitors that block the action of a particular target



David Szymkowski says drug failures can be due to poor target validation.

protein. But the only way to be completely certain that a protein is instrumental in a given disease is to test the idea in humans. Obviously such clinical trials cannot be used for initial drug development, which means that a potential target must undergo a validation process — its role in disease

must be clearly defined before drugs are sought that act against it, or before it is used to screen large numbers of compounds for drug activity.

Deciding to develop a drug against a particular target is a big commitment in terms of time and money. Once a target enters a pharmaceutical company's pipeline, it can take about 12 years to develop a marketable drug. Each new drug that reaches the market represents research and development costs of close to US\$1 billion.

"Reducing failures early in development is far more important than filling a pipeline with poorly chosen late-stage products likely to fail, and fail expensively," says Szymkowski.

Model interactions

Computer models are a fast, relatively cheap option for initial screening of both targets and potential drugs. These models usually focus on how the two types of candidate structures interact with each other.

De Novo Pharmaceuticals in Cambridge, UK, has a suite of software for the process, covering virtual screens,

A QUESTION OF FORM

For any protein, it is not guaranteed that all of its isoforms will have the same function, so it is important to work out which forms are valid drug targets. Many proteins are expressed as different isoforms as a result of alternative splicing of the precursor messenger RNAs (mRNAs) and post-translational modifications. Techniques such as gene knockouts effectively remove all isoforms, but manipulation of mRNA with antisense technology or RNA interference (see 'The silent treatment', page 343) cannot distinguish between isoforms that differ in their post-translational modifications.



Transcription factor production at Sangamo.

To overcome this problem, biotechnology company Sangamo BioSciences in Richmond, California, has developed a system that allows the expression of endogenous genes to be altered in cells or whole-animal models. It uses zinc-finger transcription factors in which the DNA-recognition domain is coupled to a functional domain that allows the expression of the target gene to be up- or downregulated. The advantage of this method over gene knockouts or transgenes is that the expression levels of all the isoforms of an endogenous target gene can be specifically manipulated.

This technique has been used by a group including researchers at Sangamo and Frank Giordano at Yale University to investigate the growth of blood vessels by activating the endogenous expression of vascular endothelial growth factor (VEGF). The team found that all isoforms of VEGF are needed to stimulate the growth of blood vessels that do not leak (E. J. Rebar *et al. Nature Med.* **8**, 1427–1432; 2002).

Casey Case of Sangamo is excited by what he sees as the growing awareness of the biological importance of alternative mRNA splicing. "Conventional overexpression methods for target validation miss the biological consequences of this important phenomenon," he says.

Another approach to the problem has been developed by Xerion Pharmaceuticals in Martinsried, Germany. It uses a system that can remove specific proteins by laser inactivation rather than through genetic manipulation. This route to studying protein function is more direct than genetic means, and may circumvent potential artefacts of overexpression; but, unlike genetic manipulations, Xerion's methods are still limited to cells in culture.

C.S.

docking programs and ligand-based design.

If the structure of the target protein's active site is known, the company's SiteExplorer can predict potential drug-binding sites, and can evaluate interactions between these sites and the drug candidates. If the structure is unknown, then its Quasi2 software will produce a virtual protein based on molecular features known to be important in binding in other targets. Drugs can then be designed against the model.

De Novo also offers software to aid the design of chemical probes used in target validation. The SkelGen suite of programs can then use these data to generate new chemical structures optimized for interaction with a target's active site.

The company is collaborating with GeneFormatics of San Diego, California, in a programme focused on inhibitors of the M10 family of matrix metalloproteinases, enzymes that are involved in cancer and inflammatory disorders. GeneFormatics is using proteomics to identify the target enzymes and characterize their active sites, while De Novo is running docking models and virtual screens of small molecules against the proteins.

Software that can model drug-receptor interactions is available from a number of companies including Tripos of St Louis, Missouri; Accelrys in San Diego, California; and Metaphorics in Mission Viejo, California. In addition, some software is available free to researchers

at non-profit organizations, such as AutoDock 3 made by the Scripps Research Institute in La Jolla, California, and GOLD from the Cambridge Crystallographic Data Centre, UK. Molsoft in La Jolla, California, which makes the ICM molecular modelling software, last month released an ICM browser for the Apple Macintosh.

The Accelrys suite of structural homology programs identifies the possible function, fold family and three-dimensional structure of target proteins by comparing them with sequences and structural homologues of known function. Once the protein's structure is determined, functional information can be gleaned using different modules within Accelrys's Insight II program, which supports a number of processes including X-ray crystallography, nuclear-magnetic-resonance studies and protein engineering.

Target Engine from LION Bioscience in Heidelberg, Germany, aids target selection by offering the ability to analyse gene sequence and expression data, find homologous structures, map potential functional features onto protein structure, view related gene annotation and protein pathway information and use text mining to find functional relationships.

In biotherapeutics, proteins themselves are developed as active drugs. One software suite designed to help optimize protein function is Protein Design Automation (PDA) produced by Xencor. "We don't



Purifying proteins for functional assessment at Xencor

screen DNA sequences," says Szymkowski. "More specifically, PDA computationally screens massive numbers of amino-acid changes in a known protein structure." It then derives functional information from the three-dimensional protein structure and designs novel features into the protein to optimize its function.

Sense reversal

Another route to target validation hinges on disrupting gene expression to reduce the amount of the corresponding protein, and so identify the physiological role of the target. Examples of this technique include gene knockouts, antisense

THE SILENT TREATMENT

MARLENA EMMONS

RNA interference (RNAi) is a selective method of silencing protein expression at the post-transcriptional level. Double-stranded RNA specific to the gene to be silenced is introduced into the cell, where it is processed into short single-stranded RNA fragments. Antisense strands, complementary to the fragments, bind to the target RNA and prompt enzymes to disable it. The effect is to destroy all the target messenger RNA, effectively halting production of the protein.

Genetica, a start-up company in Cambridge, Massachusetts, co-founded by Gregory Hannon at Cold Spring Harbor Laboratory, New York, is developing RNAi technology for high-throughput target validation in mammalian cells. Hannon and his colleagues in New York have done much of the recent work in the stable RNAi suppression of gene expression in mammalian cells.

Scientists at Bristol-Myers Squibb, based in Princeton, New Jersey, are trying to create RNAi reagents for the entire proteome, to allow analysis of all expressed genes. But there are still problems with the method, says Pam Carroll, senior research investigator in applied genomics at the company. "Most researchers use synthetic RNA double-stranded oligos that are expensive and there is still variability in response. Nonetheless, it has been an amazing year for RNA interference as



Gregory Hannon has established RNA interference in mammalian cells.

a mammalian target validation technology," she says. For example, only recently has RNAi been successfully used in mammalian model systems.

Another exciting development in RNAi is the use of small interfering RNA (siRNA) compounds. These are "efficient, specific and relatively non-toxic", says Dmitry Samarsky, manager of technology development at functional-genomics firm Sequitur in Natick, Massachusetts. But he notes that getting these compounds into cells is still a challenge. For researchers interested in a convenient approach to RNAi, companies such as OligoEngine in Seattle, Washington, and Ambion in Austin, Texas, offer siRNA kits and expression vectors, as well as custom siRNA synthesis. Custom siRNAs are also now supplied by many RNA companies such as Dharmacon in Lafayette, Colorado; Prolog in Hamburg, Germany; and QIAGEN in Venlo, the Netherlands.

The biotech firm Benitec, in St Lucia, Australia, has circumvented the difficulties of introducing double-stranded RNA into cells by developing a DNA template that produces double-stranded hairpin-loop molecules within the cell, which trigger RNAi. Last September, Benitec announced that it had made the first transgenic mice in which targeted endogenous genes were suppressed using RNAi.

C.S.

technology and RNA interference (RNAi).

In the realm of drug discovery, antisense technology — the use of short oligonucleotides to target specific messenger RNAs for destruction — was developed as a way of finding oligonucleotide-based drugs that interfere with gene expression, rather than with protein function. But the technology is currently enjoying greater success as a high-throughput method of target validation because it offers a highly specific and efficient way to inhibit the expression of potential target proteins *in vitro* and *in vivo*.

GeneTrove, the genomics division of Isis Pharmaceuticals in Carlsbad, California, is one of the companies active in this field. It is focusing on the untapped pool of potential therapeutic target RNAs for both target validation and drug discovery, says Nicholas Dean, GeneTrove's vice-president of functional genomics. It offers custom target-validation packages that include optimized antisense inhibitors against any target of interest and control oligonucleotides for testing in cell-culture model systems. It also applies antisense technology to target validation *in vivo* in animal models.

Biognostik, a biotechnology company in Göttingen, Germany, offers a drug-target validation kit that can be used *in vitro* or *in vivo*. It includes five target-specific phosphorothioate antisense inhibitors and two random-sequence oligonucleotides to control for nonspecific



GeneTrove analyses the effects of antisense on gene expression.

effects. It has also developed a sequence-design system called RADAR, which determines antisense oligonucleotides based on specificity, minimal nonspecific effects or protein binding, and the ability to be taken up into cells.

Sequitur, a functional-genomics company in Natick, Massachusetts, has a slightly different approach to rapid target validation. It combines an antisense library with high-throughput DNA microarray assays to test the effects of the antisense molecules on gene expression. The company's technology was used recently to validate a major therapeutic target for Alzheimer's disease. Sequitur also carries out target validation based on RNAi (see 'The silent treatment', page 343).

Custom phosphorothioate antisense oligonucleotides for research are available

from firms such as Sigma-Genosys at the Woodlands, Texas; atugen in Berlin, Germany; and Integrated DNA Technologies in Coralville, Iowa. Gene Tools in Philomath, Oregon, offers morpholino antisense oligonucleotides, and Danish companies Cureon in Copenhagen and Exiqon in Vedbæk offer modified oligonucleotides based on 'locked nucleic acid' technology that can be used for antisense.

The proteomics approach

One disadvantage of doing target validation at the genetic level is that many genes produce several different protein isoforms, which can have subtly different functions (see 'A question of form', page 341). Post-translational modifications can also give protein variations. As a result, a developing approach in target validation is to focus on manipulating the activity of the potential target protein itself. "As the vast majority of drugs target proteins, validating targets is best done by modulating protein activity, not expression levels," Szymkowski says.

Proteomics — the study and manipulation of the protein make-up of a cell — is making it easier to distinguish and target just one specific form of a protein. This allows researchers to avoid unwanted changes in the expression of other proteins — another potential drawback of genetic manipulations.

Stefan Henning, director of functional biology at Xerion Pharmaceuticals in

A WHOLE PICTURE

Beyond Genomics in Waltham, Massachusetts, is taking a broad view for target validation — its approach hinges on systems biology. The company first characterizes potential targets within their normal biochemical pathways. "We measure cells, tissues and body fluids at multiple bioanalytical levels, including transcript, protein, and endogenous metabolite and enzyme product," says Aram Adourian, the company's senior director of advanced technologies and strategic development.

It then compares these profiles with similar profiles from disease states, and correlates features of the profile with the disease in question using pattern-recognition algorithms and bioinformatics techniques. The result is a set of potential targets, which can then be perturbed by RNAi or transgenics, for example. The subsequent biological changes are monitored by the same systems-biology approach.

As a target-validation tool, systems biology is "a truly exciting method for illuminating causative relations between target modulation and effects on disease", says Adourian. By learning more about the biological pathways in which the target is involved, it is possible to predict the potential toxicity of perturbing its normal function, he says.

Adourian believes that an integrative approach to target validation is more likely to yield the highest-quality targets: "One ultimately needs to examine the system as a whole, with its interacting networks and components of genes, proteins, metabolites and other molecules, to begin to assemble a unified and contextualized perspective on the role, function and relevance of a potential target."

Another firm taking a wide view of target validation is MDS Proteomics in Toronto, Canada, which uses bioinformatics, gene-expression information, high-throughput protein analysis and protein-pathway biology to identify prospective targets and gain all-round evidence of their cellular role.

C.S.



Combining many techniques is key to systems biology.

Martinsried, Germany, agrees that validation at the proteomics level is a powerful approach. “On a technical level, the development of protein microarrays, multidimensional liquid-based protein separation and technologies that manipulate protein expression and protein–protein interactions will have their impact,” he says.

Xencor has developed ProCode, which enables researchers to study the functions of a cell’s protein make-up. A ProCode library is a protein-expression library from any cell or tissue of interest, in which every protein (after translation) is tagged with a plasmid, a small circular piece of DNA containing its corresponding complementary DNA (cDNA). The library can be expressed in cultures of the appropriate mammalian cells so that proper protein folding and processing are retained. The expressed proteins can be screened for their interactions with potential drugs, and the cDNA tags allow easy identification of any protein that gives a positive reaction.

Xerion’s XCALibur carries out simultaneous identification and functional validation of potential drug targets. Using target-specific antibodies to identify the proteins and chromophore-assisted laser inactivation (CALI) to ‘switch off’ target proteins by photochemically modifying their functional sites, XCALibur can validate specific targets for particular diseases or find new potential targets with disease-associated functions.

XCALibur is incorporated into the Xstream platform, which takes a disease-based approach to target validation. It searches for hits from a suite of antibodies specifically created against the proteome of a diseased cell. The antibodies bind near to functional sites of proteins and contain dyes that are released by CALI, thereby inactivating the proteins’ functional sites. If this inactivation has an effect on the disease, the protein is precipitated by the attached antibody and analysed by mass spectroscopy and database searches.

Validation *in vivo*

One of the most important tests for a potential drug is an assessment of its role in disease in an animal model. But animal models for certain diseases, such as psychiatric illnesses, are extremely difficult to develop.

“The greatest challenge in target validation is the procurement or development of the correct animal models for the human disease in question,” says Bob Gordon, De Novo’s vice-president of biology. “For example, there are few, if any, reliable animal models for stroke. So validation is effectively done in phase III trials in the clinic. Progress in this disease



Xerion’s XCALIBUR switches off target proteins by laser.

area is understandably slow and expensive.”

In vivo target validation using gene knockouts, in which genes are deleted or disrupted to halt their expression, is a powerful method of predicting drug action. “Many of the targets for the top-selling drugs of the biopharmaceutical industry have been knocked out,” says Arthur Sands, president and chief executive of Lexicon Genetics in the Woodlands, Texas.

This kind of target validation is based on the assumption that knocking out the gene for the potential target has the same effect as administering a highly specific inhibitor of the target protein *in vivo*.

“With the effective use of mouse knockout technology, expensive drug-discovery activities can be focused on the drug targets that are most likely to lead to breakthrough therapeutics,” says Sands. Furthermore, target-specific side effects can be discovered before time and money are invested in drug design.

But mammals are not the only creatures in use — zebrafish have recently entered the fray as a model animal for some human diseases. The fish are more affordable, easier to keep, and faster to raise than mammals, giving a higher-throughput system. Drugs can also be tested for toxicity and their potential therapeutic activity against the target more easily than in mammals.

Perhaps surprisingly, genes that cause disease in zebrafish are similar to those in humans, for example in angiogenesis, inflammation and insulin regulation. The transparency of zebrafish embryos also makes them suitable for large-scale, high-throughput genetic and drug screens.

Zygogen in Atlanta, Georgia, has developed a transgenic zebrafish system called Z-Tag which can be used for target validation. The company can also make

various zebrafish organs visible by tagging the tissues with fluorescent markers.

One of the most widely used models of human disease is the mouse, but working with mice can be both time-consuming and expensive. Lexicon Genetics has met this challenge by industrializing the generation of mouse knockouts, using gene targeting, gene trapping and mouse embryonic-stem-cell technologies. The result is the company’s Genome 5000 programme, which aims to analyse 5,000 genes over the next five years — over 750 have already been done. Custom transgenic and knockout mice are also available from Deltagen in Redwood City, California, and memorec stoffel in Cologne, Germany.

Researchers are slowly but surely making progress in validating the targets revealed by the Human Genome Project. But proving that a target protein has a causative role in human disease remains a real challenge. “The most exciting technologies are, and will be, those that address the issue of elucidating causative roles of targets in human disease, as opposed to simple associations,” says Aram Adourian, a senior director at the biopharmaceutical firm Beyond Genomics in Waltham, Massachusetts (see ‘A whole picture’, page 345).

More challenging tasks lie in discovering the effects of interactions between newly validated targets in both healthy and diseased models. Such complex information will require not only information systems to correlate multiple variables and outcomes, but also a sophisticated knowledge of protein–protein interactions under a variety of conditions. But for now, researchers are taking it one target at a time.

Caitlin Smith is a science writer based in Portland, Oregon.

Company	Products/activity	Location	URL
Tools for <i>in silico</i> and structure-based target validation			
Accelrys	Software for sequence and genome analysis, protein-structure analysis and homology modelling, data management	San Diego, California	www.accelrys.com
Chemical Diversity Labs	Orphan GPCR- and orphan nuclear-receptor-targeted libraries of small molecules	San Diego, California	www.chemdiv.com
Compugen	GenCarta annotated human genomic, transcriptome and proteome database; computational techniques for drug discovery	Tel-Aviv, Israel	www.cgen.com
De Novo Pharmaceuticals	Structure-based target selection and analysis	Cambridge, UK	www.denovopharma.com
Entelos	PhysioLab large-scale <i>in silico</i> models of human diseases	Menlo Park, California	www.entelos.com
Metaphorics	DockIt flexible ligand-docking program; Luna database of protein-ligand structural information	Mission Viejo, California	www.metaphorics.com
Molsoft	Molecular-modelling software	La Jolla, California	www.molsoft.com
Structural GenomiX	Automated high-throughput protein crystallography for drug discovery	San Diego, California	www.stromix.com/
Syrrx	Automated high-throughput protein crystallography for drug discovery	San Diego, California	www.syrrx.com
Tripes	SYBYL software for protein-structure prediction and modelling, FlexX ligand-docking program, bioinformatics software	St Louis, Missouri	www.tripos.com
Xencor	PDA technology for computational protein design	Monrovia, California	www.xencor.com
Proteomics			
Beyond Genomics	Target validation using a systems-biology approach	Waltham, Massachusetts	www.beyondgenomics.com
Biacore	Analysis and measurement of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com
Hamilton Life Science Robotics	MICROLAB workstations for protein crystallization, MALDI target spotting	Bonaduz, Switzerland	www.hamiltoncompany.com
Inpharmatica	Biopendium and CeleraEdition Biopendium proteome annotation resources; PharmaCarta	London, UK	www.inpharmatica.com
Jerini Array Technology	Substrate identification service for kinases using pepSTAR peptide microarray technology	Berlin, Germany	www.jerini.com
KeyNeurotek	Functionality and tissue-based screening for target identification, validation and compound/drug testing; disease models for neurodegeneration and ischaemia/stroke	Magdeburg, Germany	www.keyneurotek.de
LI-COR Biosciences	Infrared imaging system for western blot and in-cell western blot assay analysis	Lincoln, Nebraska	www.licor.com
MDS Proteomics	Target validation using proteomics-based systems biology	Toronto, Canada	www.mdsproteomics.com
Tecan	Free-flow electrophoresis, protein digest and crystallization systems	Mannedorf, Switzerland	www.tecan.com
Zyomyx	Human cytokine-profiling biochip system	Hayward, California	www.zyomyx.com
Genomics			
Ardais	BIGR library of biomaterials and information for genomics-based biomedical research	Lexington, Massachusetts	www.ardais.com
Galapagos Genomics	PhenoSelect adenovirus vector/transgene system for target validation	Leiden, the Netherlands	www.galapagosgenomics.com
GeneFormatics	Genomics-based drug target discovery	San Diego, California	www.geneformatics.com
Incyte Genomics	Annotated gene and expressed sequence tag databases; Proteome BioKnowledge Library species-specific protein information databases; bioinformatics software	Palo Alto, California	www.incyte.com
Sangamo BioSciences	Zinc-finger transcription factors for gene regulation	Richmond, California	www.sangamo.com
Hybridon	Cyclicons synthetic fluorescent DNA probes for target validation, drug discovery and PCR-based gene amplification	Cambridge, Massachusetts	www.hybridon.com
Target validation services			
Althea Technologies	PCR-based high-throughput gene-expression profiling	San Diego, California	www.altheatech.com
Biovex	Vaccine development; target validation through gene-delivery system	Oxford, UK	www.biovex.com
Eurogentec Proteomics	Validation of <i>in silico</i> and genomically derived targets	Berlin, Germany	www.vita-proteomics.com
Genovac	Functional antibodies against GPCRs	Freiburg, Germany	www.genovac.com
Genzyme Molecular Oncology	<i>In vitro</i> and <i>in vivo</i> target validation services	Framingham, Massachusetts	www.genzymemolecularoncology.com
Hybrigenics	<i>In silico</i> prevalidation and experimental validation services	Paris, France	www.hybrigenics.com
Immusol	Ribozyme-based Inverse Genomics platform for identification of drug target genes	San Diego, California	www.immusol.com
Phenomix	Identification of disease models, drug targets and validation	La Jolla, California	www.phenomixcorp.com
Proteome Factory	Target validation <i>in vitro</i> and <i>in vivo</i>	Berlin, Germany	www.proteome-factory.de
ProQinase	Technologies to validate novel protein kinases as drug targets in oncology	Freiburg, Germany	www.proqinase.com
Xerion Pharmaceuticals	XCALibur antibody-based target-inhibition technology	Martinsried, Germany	www.xerion-pharma.com
Zeptosens	Array-based detection of biomolecules	Witterswil, Switzerland	www.zeptosens.com
Software and bioinformatics			
BioRad	WorksBase bioinformatics software for proteomics	Hercules, California	www.discover.bio-rad.com
Celera	Discovery System web-based informatics tools for accessing the Celera annotated genomes databases; bioinformatics services; DNA sequencing; target validation	Rockville, Maryland	www.celera.com
Informax	Desktop bioinformatics software for sequence and genome analysis, microarray gene-expression data; contract bioinformatics services	Bethesda, Maryland	www.informaxinc.com
Inpharmatica	Biopendium and CeleraEdition Biopendium proteome annotation resources; PharmaCarta	London, UK	www.inpharmatica.com
LifeSpan BioSciences	Proprietary gene-expression and protein-localization databases and data-mining tools for drug-target validation and drug discovery	Seattle, Washington	www.lsbio.com
LION Bioscience	Software for sequence and genome analysis, Target Engine for drug target selection and Lead Engine for database development and management	Heidelberg, Germany	www.lionbioscience.com
Proteomic Solutions	Melanie software for two-dimensional gel analysis; distributor for Genomics Solutions	St Marcel, France	www.proteomicsolutions.fr
Rosetta Biosoftware	Rosetta Resolver and Rosetta Luminator analysis system for gene-expression data	Kirkland, Washington	www.rosettatabio.com
Molecular Biology Insights	Oligo program to identify oligonucleotides for PCR, DNA sequencing and other applications	Cascade, Colorado	www.oligo.net

Company	Products/activity	Location	URL
Transgenics			
Deltagen	Gene knockout and gene function databases; transgenic mice for target validation	Redwood City, California	www.deltagen.com
Lexicon Genetics	Gene knockout and gene function databases; transgenic mice for target validation	The Woodlands, Texas	www.lexgen.com
Ingenotyping	Custom transgenic mouse models	Martinsried, Germany	www.ingenotyping.com
memorec stoffel	Transgenic and knockout mice for target validation; cDNA microarrays; SAGE analysis	Cologne, Germany	www.memorec.com
OzGene	Transgenic and knockout mice for target validation	Canning Vale, Australia	www.ozgene.com
Zygogen	Transgenic zebrafish system for target validation	Atlanta, Georgia	www.zygogen.com
Products and services for antisense and RNA interference			
Ambion	Silencer range of kits, vectors, siRNAs and reagents for RNAi; kits and products for RNA synthesis, isolation, quantitation and analysis	Austin, Texas	www.ambion.com
atugen	Custom phosphorothioate antisense RNAs for drug-target validation	Berlin, Germany	www.atugen.com
Benitec	Proprietary gene silencing RNAi technology	St Lucia, Australia	www.benitec.com.au
Biognostik	Antisense oligonucleotide custom synthesis kits and services; ANTISENSE drug-target validation kits for <i>in vivo</i> or <i>in vitro</i> studies	Göttingen, Germany	www.biognostik.com
Cureon	Custom antisense locked nucleic acid oligonucleotides	Copenhagen, Denmark	www.cureon.com
Cytogenix	Target validation using antisense RNA technology	Houston, Texas	www.cytogenix.com
Dharmacon	siRNA kits and arrays; custom RNAs; large-scale RNA synthesis	Lafayette, Colorado	www.dharmacon.com
Eurogentec	Custom siRNA for RNAi	Seraing, Belgium	www.eurogentec.be
Exiqon	Microarray slides and chips; LNA oligonucleotides	Vedbæk, Denmark	www.exiqon.com
ExpressOn BioSystems	Scalable RNA structure mapping for design of antisense and optimization of RNAi	Roslin, UK	www.expresson.co.uk
Gene Tools	Morpholino antisense oligonucleotides	Philomath, Oregon	www.gene-tools.com
GeneTrove	Antisense technology for target validation	Carlsbad, California	www.genetrove.com
Imagenex	Plasmid-based GeneSuppressor siRNA kits for RNA interference	San Diego, California	www.imgenex.com
Integrated DNA Technologies	Custom oligonucleotides including phosphorothioate antisense oligonucleotides	Coralville, Iowa	www.idtdna.com
Invitrogen	Kits and reagents for genomics, proteomics, molecular and cell biology; Lipofectamine 2000 for siRNA transfection	Carlsbad, California	www.invitrogen.com
InvivoGen	siRNA cloning vectors, custom siRNA in plasmid and viral vectors	San Diego, California	www.invivogen.com
Mirus	Products for gene transfer and RNAi	Madison, Wisconsin	www.genetransfer.com
MWVG Biotech	Custom oligonucleotide synthesis; Dharmacon siRNA kits and arrays; custom sequencing; microarrays and oligo sets	Ebersberg, Germany	www.mwvg-biotech.com
New England Biolabs	HiScribe RNAi transcription kit	Beverly, Massachusetts	www.neb.com
Novagen	RiboJuice siRNA transfection reagent	Madison, Wisconsin	www.novagen.com
OligoEngine	pSUPER RNAi vector system for gene silencing; custom oligonucleotides	Seattle, Washington	www.oligoengine.com
Proligo	Custom and speciality DNA and RNA oligos; siRNA oligos for RNAi	Hamburg, Germany	www.proligo.com
QIAGEN	Custom siRNA and transfection reagents; Cancer siRNA Oligo Set	Venlo, The Netherlands	www.qiagen.com
ResGen	Custom antisense and phosphorothioate antisense RNAs; GeneStorm expression-ready full-length clones	Huntsville, Alabama	www.resgen.com
Sequitur	Antisense and siRNA compounds for target validation	Natick, Massachusetts	www.sequiturinc.com
Sigma-Genosys	Tools for molecular biology; phosphorothioate antisense oligonucleotides	The Woodlands, Texas	www.sigma-genosys.com
SomaGenics	Target-validation services based on RNA Lasso antisense technology	Santa Cruz, California	www.somagenics.com
Spring Bioscience	Kits for siRNAs; gene cloning, target screening and protein-expression analysis products	Fremont, California	www.springbio.com
Stratagene	GeneEraser siRNA transfection reagents	La Jolla, California	www.stratagene.com
Upstate	KinaseProfiler specificity testing, IC50Profiler Express and services for crystallography assay development, bulk protein production and purification	Charlottesville, Virginia	www.upstate.com
General			
Affibody	High-throughput protein identification and characterization by binding proteins	Stockholm, Sweden	www.affibody.com
Agilent	Workstation for RNA, DNA or protein analysis; instruments for genomics and proteomics	Palo Alto, California	www.agilent.com
Amersham Biosciences	IN Cell Analyser 3000 confocal imaging system for rapid cellular assays	Little Chalfont, UK	www.amershambiosciences.com
Apogent Technologies	Labware and equipment for life sciences	Portsmouth, New Hampshire	www.apogent.com
Applied Biosystems	DNA sequencers, protein sequencers, mass spectrometers, chromatography systems, PCR systems, peptide synthesizers, nucleic-acid synthesizers	Foster City, California	home.appliedbiosystems.com
Chromagen	Fluorescence-based products, kits and services for gene expression and protein detection and assay	San Diego, California	www.chromagen.com
Ciphergen	ProteinChip system for protein identification and characterization from complex mixtures using SELDI technology	Fremont, California	www.ciphergen.com
Cytomix	One- and two-dimensional gel analysis, mass spectroscopy and amino-terminal sequencing for proteins, cDNA cloning, DNA sequencing, expression profiling, recombinant protein expression, cell-line production	Cambridge, UK	www.cytomix.com
Greiner Bio-One	Consumables for high-throughput research and diagnostics, microplates	Frickenhausen, Germany	www.greinerbioone.com
LumiCyte	Protein biomarker profiles for drug development	Fremont, California	www.lumicyte.com
PanVera	Protein reagents and assays for receptor and enzyme activity	Madison, Wisconsin	www.panvera.com
Proteodyne	BioCube automated systems for research and genome-based drug discovery	Windsor, Connecticut	www.proteodyne.com
Roche Diagnostics	Reagents and kits for molecular biology, functional genomics and proteomics research	Lewes, UK	www.roche-applied-science.com
Thermo Finnigan	Instruments for chromatography and mass spectroscopy; ProteomeX workstation for automated LC/MS protein analysis	San Jose, California	www.thermo.com

● See advertisement

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe
e-mail: h.watanabe@naturejp.com

naturejobs

The glory of autonomy

To paraphrase John Donne, no institution is an island. But Salvador Moncada, director of the Wolfson Institute for Biomedical Research in London, feels that cultivating self-sufficiency can be a virtue. The independently funded institute opened on the University College London campus in 1997. Three years later it moved into its own building, a red-brick Victorian edifice. It has since set up its own technology-transfer office and created four companies that between them have raised £100 million (US\$160 million). Now, the institute is taking another step forwards: it is spinning off its own contract-research division.

All of these moves have meant job creation and training, particularly in medicinal chemistry. "The need for medicinal chemists is so big, we've got people knocking on our doors morning, noon and night," says Moncada. The contract-research division is helping the drug-development industry, providing funds to fuel the expansion of the institute's medicinal-chemistry department from 20 to 30 people, and offering more training opportunities for students and postdocs.

Moncada says he has found that the more the institute exerts its autonomy, the greater the need for its researchers to work with each other. Ironically, the building that houses them, a former hospital, was designed as a cruciform supposedly to avoid infectious diseases from travelling from wing to wing. "That's quite a problem if you want to create interactions," says Moncada. So to encourage intellectual cross-pollination, he has had to do some social engineering, which includes providing "the best coffee around" and placing chemists doing pharmacology next to geneticists doing microarrays.

Such moves ensure that the Wolfson Institute is an island that people will want to journey to, as well as communicate with.

Paul Smaglik
Naturejobs editor



Contents

CAREERS & RECRUITMENT

A permanence problem in
the United States p354

Europe's struggle to
strengthen its science base p356

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Victims of success

The doubling of public funds for life-science research in the United States has increased the number of postdocs, but it has yet to create significantly more permanent academic positions, says Eugene Russo.

As the economics of science in the United States changes, so too does the scientific enterprise. There are more authors per paper, more graduate schools, more postdocs. Yet there are relatively few entry-level faculty positions. Meanwhile, in areas such as biomedical research, large-scale investment continues to increase significantly.

Nowhere is the bittersweet mix of bulging budgets and postdoc proliferation more apparent than in the life sciences, where the prospective investigator faces an often frustrating journey. A serious, system-wide post-postdoc bottleneck is changing the composition of temporary, permanent and student workers in the lab.

"The operating model of science has changed," says Elias Zerhouni, director of the National Institutes of Health (NIH). "It's not just an NIH issue. I think it's really a systemic issue." As research teams grow larger and require more complex skill sets, they need more postdocs and graduate students, who in turn take longer to complete training.

A couple of decades ago, prospective scientists could move easily from one lab to another. Zerhouni notes that scientists such as James Watson, co-discoverer of DNA's double-helix structure, and Marshall Nirenberg, who helped to decipher the genetic code, made major discoveries while still in their 20s and 30s — in today's terms, both might have received their Nobel prizes before they received their first NIH grant. "What I always used to say is you don't become a thoroughbred by longevity," Zerhouni says.

Part of the problem, he suggests, is that the system mainly recognizes those scientists who succeed in the well-trodden career pathway towards being a faculty member. Finding and promoting additional pathways may be necessary, says Zerhouni. "I look at it as probably one of the most important long-term priorities for the NIH director to focus his energies on," he says, adding that he has started to "put a process in place", although he hesitates to discuss specific solutions as yet.

Other scientists, economists and policy-makers have also offered analyses and possible remedies. Reports from the National Academy of Sciences (NAS) in 1994, 1998 and 2000 recognized the flaws in the system,



Elias Zerhouni wants to promote fresh pathways to scientific success.

particularly the plight of postdocs seeking their first faculty positions — the average first-time grantee is now over 35 years old, largely because of a longer training period. "I haven't seen anything to say that those trends have changed," says Walter Schaffer, acting director of the NIH's office of extramural programmes.

TREND SETTERS

The 1998 NAS committee attempted to offer solutions. Universities and researcher institutions, said its report, should not continue to expand enrolment in graduate programmes or develop new ones unless they are directed at a specific need (for example, in a burgeoning subdiscipline such as bioinformatics). It was this recommendation that "got the most heat", says the committee's chair, Princeton University president Shirley Tilghman, because it challenged the status quo.

The report in 1998 also advised federal agencies to fund more graduate students through training grants — which allow better scrutiny of the quality of training offered — rather than through principal investigators' research grants. The funding of graduate students and postdocs with research grants has caused the number of trainees in the lab to expand out of proportion with the number of permanent jobs they subsequently seek, says Tilghman. The trend has become more pronounced over the past five years, as successive US budgets have seen the NIH's budget double from its 1998 level. According to Schaffer, the number of students — about 17,000 — supported by training grants and fellowships has changed little during the past decade.

Also in 1998, the American Society for Cell Biology

Shirley Tilghman (right) believes the balance between postdocs and the number of permanent positions in the United States needs to be redressed.



Making the transition

Competition among the growing throng of biomedical research postdocs for a first faculty position is tough. But there are a few funding mechanisms that can separate the exceptional candidates from the crowd.

As part of a pilot programme, many institutes under the umbrella of the US National Institutes of Health now offer a handful of transition awards. These provide additional funding during awardees' postdoc fellowships, and offer a promise of start-up funds for awardees' first faculty position. Most are intended to forge a transition from intramural labs to extramural academic institutions.

This pledge of financial support is a major hiring incentive for the prospective employer. Although still postdocs, the awardees get several thousand dollars over and above their standard salary. Once the postdocs become faculty members at an extramural institution, they can receive up to \$175,000 per year for three years, roughly half of which can be put towards salary.

The Burroughs Wellcome Fund has similar programmes for biomedical and interdisciplinary sciences, although most of its grants have been put on hold due to the strains the sagging stock market has placed on many foundations. **E.R.**

conducted a survey of over 2,400 members at various points in their careers. Later, the society, in collaboration with Harvard economist Richard Freeman, did 'life histories' of some 30 well-established labs. They obtained some discouraging results.

Older scientists reported that it is now harder for young scientists to secure a faculty position. Many postdocs complained that today's emphasis is on research productivity over training and that their work differs little from that of graduate students. "Postdocs said: 'I don't feel like I'm in training, I feel like I'm a pair of hands,'" recalls co-author Frank Solomon, a biology professor at the Massachusetts Institute of Technology. He also found that prospective postgrads sought the advice of dissatisfied postdocs and graduate students, suggesting, he says, that a dissatisfied group of people might "nucleate greater dissatisfaction".

Just over a year ago, a research network of labour economists was formed to discuss analyses of the science and engineering labour markets, sponsored by the Alfred P. Sloan Foundation. Ideally, says Freeman, it will offer independent analyses of government, academic and industry institutions, and will share its information with professional societies and policy-makers. A recent report, he says, shows that, as in the United States in general, the rich have got richer and the poor poorer: although science has become more collaborative and research groups have grown, economic inequality among team members has increased.

Tuning the US engine of biomedical science will be tricky. The NIH has increased stipends, not only to raise postdocs' pay but to help stem their proliferation

by increasing their cost to a principal investigator. The agency has also started a 'transition award' programme that helps cream-of-the-crop postdocs secure their first faculty position with start-up funding (see 'Making the transition', left).

Some universities, encouraged by postdoctoral groups — including the National Postdoctoral Association, which met for the first time this month — have also taken action. Stanford's medical school, which has three-quarters of the university's postdocs, has been one of the most proactive. Last September, it raised the minimum postdoc pay to \$35,000, limited fellowships to five years (at which point the postdoc leaves or becomes a permanent employee), improved health benefits, offered more subsidized housing and instituted a mentoring programme to emphasize the training element of the postdoctoral experience. The agreement arose from "a really good dialogue" among faculty members, the postdocs and the administration, says William Nelson, the medical school's senior associate dean of graduate and postdoctoral education.

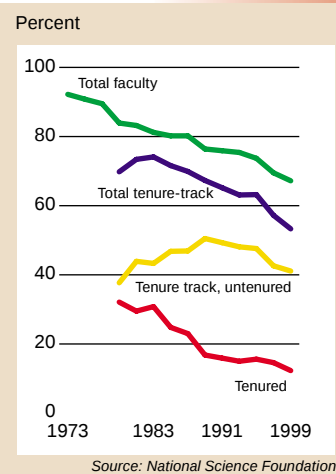
"Some faculty members have complained loudly," admits Nelson. Principal investigators in the middle of multiyear grants were particularly displeased by the unexpected funding changes; postdoc-heavy chemistry labs were hit hard. "My comment to that has been: 'I'm sorry but we have to do this somehow,'" Nelson says.

But improving the postdoc situation may be the easy part. "A lot of campuses are trying to address those issues," says Paula Stephan, a labour economist at Georgia State University and a committee member for the 1998 NAS report. "What I don't see people trying to address is the harder problem, and that's what your job prospects are when you leave your postdoc."

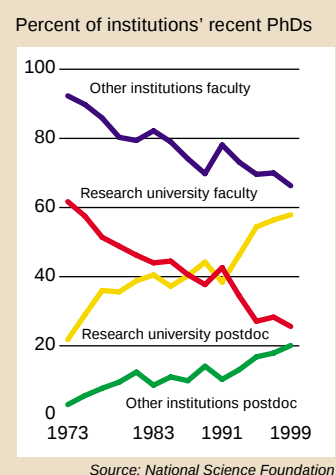
Solomon and others suggest establishing a new class of permanent bench-scientist employees and raising the bar for prospective PhD students. "I'm a big believer in questioning the role of the postdoc position," says Solomon. Perhaps, he offers, the postdoc fellowship will be used for the select few as a transition to independent research or as a way to switch disciplines.

Where would these permanent bench scientists come from? Solomon expects that many students would forgo thoughts of a faculty position in favour of a standard working day, with no obligation to teach or deliver scientific talks. "There are a lot of people who would love to have jobs doing what they love to do, which is experiments," says Solomon. ■

Eugene Russo is a freelance writer based in Takoma Park, Maryland.



Status of academic science and engineering PhDs who earned their PhD 5-7 years earlier.



Science and engineering PhDs hired into faculty and postdoc positions at research universities and other academic institutions, 1973-99.

Trends in the Early Careers of Life Scientists, NRC report 1998

♦ www.nap.edu/catalog/6244.html?onpi_newsdoc091098

Addressing the Nation's Changing Needs for Biomedical and Behavioral Scientists, NRC report 2000

♦ www.nap.edu/catalog/9827.html

Clouded vision

The European Union has dreams of becoming a centralized scientific powerhouse. But first it needs to solve its brain-drain problem. Quirin Schiermeier reports.

Until the early 1990s Finland didn't really need to invest much in research and development (R&D). The country had ample natural resources and strong trading links with the East and some of the West. But its main markets dried up when the Soviet Union collapsed and the United States underwent a recession.

In response, the Finnish government decided to create a high-tech economy, and started investing heavily in R&D — especially in information technology and telecommunications. Just over a decade later, Finland's R&D spending had risen from 1.4% of gross domestic product (GDP) to 3.4%, the second highest level in the world. And with companies such as Nokia, Finland had established itself as a leading player in the telecommunications market.

Other European Union (EU) nations are aspiring to similar transformations, by trying to increase their R&D investment to 3% of GDP by 2010. But Finland's experiences show that investment is not necessarily a panacea. Despite its strong domestic science base, even Finland is losing roughly one-third of its young investigators, mainly to the United States. But very few foreign researchers come to Scandinavia, either because of language barriers, high taxes, or because the dominance of a few industries, mainly in telecommunications, limits career opportunities.

Even in the absence of a large number of foreign researchers, the Scandinavian science base performs reasonably well. But cracks are starting to appear in Europe's overall scientific aspirations. Concerns are growing in Brussels that the level and growth rates of national investment in science, particularly in the education and training of researchers, may be inadequate for Europe's future needs. Given budget cuts in many countries and reluctant private investment, the goal of making the EU the most competitive and dynamic knowledge-based economy in the world, announced in March 2000 at the Lisbon European Council, does not seem achievable in the near future, according to research analysts. The European Commission has begun a feverish benchmarking exercise,



Knut Fagri believes that scientific recruitment is reaching crisis point.

aimed at analysing the strengths and weaknesses of European research, and at improving its output and productivity.

In many respects, data from the indicators reveal worrying differences between the EU and its competitors. Total research expenditure in relation to GDP remains lower than in the United States and Japan (see chart, below left). The proportion of researchers in the workforce is lower (see chart, opposite), or is rising more slowly, than in the United States, and the workforce is older: about one-third of current researchers will retire over the next ten years.

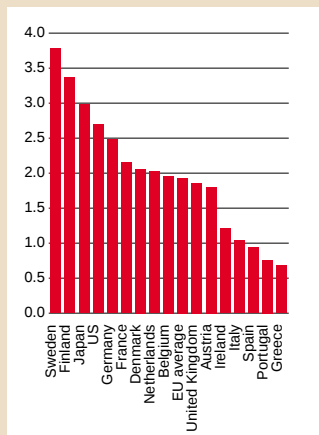
THE DRAIN GAME

Europe continues to export scientific talent to the United States. And the proportion of students entering science and technology has in some countries, including traditional science nations such as the Netherlands and Britain, dropped below 25%. In the main, only biology and computer science seem to attract a reasonably high number of students in Europe — although one exception is, again, Finland, where more than 50% of students enrol in the natural sciences.

Because many PhDs are leaving research, or leaving Europe, recruiting may soon become a "national problem", says Knut Fagri, a theoretical chemist at the University of Oslo, who chaired an EU expert group benchmarking human resources in research.

The expert group urgently recommended measures to enhance the opportunities, attractiveness and rewards of a scientific career, beginning at postdoc

Total domestic R&D spend as % of GDP



Source: DG Research

As a whole, Europe still lags behind Japan and the United States in terms of R&D spending. Figures are for 2000 except Ireland, Italy, Belgium, Netherlands, Denmark and Sweden (1999).

level. "Recently, there seems to have been a reduction in the supply of talent at this stage, as a consequence of the fact that postdoctoral opportunities are seen as leading to less attractive career paths," says the final report, published last August.

In some disciplines, the loss of newly qualified PhDs from scientific research is almost 40%, the group found. "Loss of able research talent to other nations and, much more significantly, to other activities weakens EU competitiveness," it concludes.

Loss of scientists to better-paid jobs in finance or business is a particular problem in the ten central and eastern European countries set to join the EU next year. Poland, Hungary and the Czech Republic all have a large pool of well-trained, talented young scientists who have little incentive to stay in underpaid academic research after gaining their PhDs. The same applies to Russia, whose immense science base collapsed in the early 1990s and is only slowly beginning to recover.

COMPETITIVE EDGE

So will large parts of Europe become a wasteland for science? Raoul Kneucker, former director general for international scientific affairs in the Austrian science ministry, doesn't think so. "Industry has the financial power to lure researchers away from universities," he says. "But universities throughout Europe are beginning to understand and tackle the challenges."

Recent trends in Austria and Germany, where universities are being given more autonomy and have begun to introduce market elements in recruitment policies and research activities, are encouraging signs, he believes. "This has little to do with misplaced capitalism, but a lot to do with profile building and professional management of resources," he says.

"US universities are in permanent competition with each other, looking for high-flyers to supplement their faculty and raise their scientific profile," says Martin Nowak, an Austrian biochemist and mathematician at Princeton University, New Jersey.



Europe needs to learn from the US system if it is to inspire better research, according to Martin Nowak.

"In Europe, universities tend to wait until a professor retires before they even consider new avenues in research."

Philanthropy, says Nowak, is another tradition that helps leading US universities keep their outstanding position. Only recently, he was awarded a grant, reportedly US\$30 million, by the New York financial tycoon Jeffrey Epstein, to build a new institute for mathematical biology at Harvard. The independence he can expect there would be hard to find in Europe, he says.

Nowak's long-time scientific mentor, Robert May, president of the Royal Society and former scientific adviser to the UK government, agrees that a sound "adventurous spirit" has somehow got lost. "In Europe, bureaucracy and security thinking are prevailing features of the science and innovation systems," he says. "The tolerance of failure is too low, even in the United Kingdom. Going bankrupt is still stigmatized."

But May warns against "careless comparisons". Public spending on basic science in the EU is at least equivalent to that of the United States, he says. In an analysis conducted a few years ago, May found that five of the eight main European research nations — Britain, Denmark, Sweden, Switzerland and the Netherlands — are getting more scientific value for money in terms of publications than the United States; only France, Italy and Germany performed worse (R. M. May *Science* **281**, 49–51; 1998).

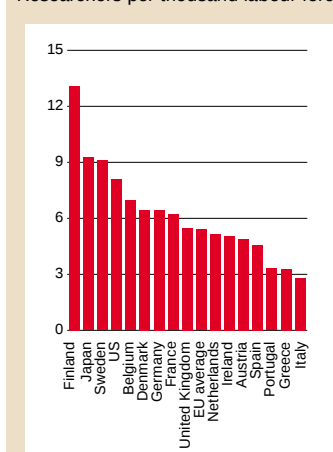
Nevertheless, several hundred thousand more academics and scientists would be needed in Germany alone if the EU is to fulfil its ambitions, according to an analysis commissioned by the German science ministry, published last month. The double dilemma, says Hariolf Grupp, a research analyst at the Fraunhofer Institute of System Analysis and Innovation Research in Karlsruhe in Germany, who coordinated the study, is that the larger EU countries have neither the necessary workforce nor enough jobs in industry and academia to back up the 3% R&D spending target.

A ray of light, says May, is that the European Commission has doubled its efforts to mobilize human capital. Under its Sixth Framework Programme (2003–06), the overall budget for Marie Curie Actions — grants aimed at improving training and mobility of researchers — has increased by 50% to E1.58 billion (US\$1.64 million; see *Nature* **421**, 296–297; 2003). The commission is also trying, with national science ministries, to remove legal and cultural obstacles to the mobility of researchers.

But will researchers go on to find work? So far, says May, little has been done to create jobs for the many young scientists that Europe so urgently needs. This has led many trainees, says one EU official, to wish that they could be Marie Curie fellows forever.

Quirin Schiermeier is *Nature's* German correspondent.

Researchers per thousand labour force



Source: DG Research

Finland has a clear lead in terms of its scientific labour force. Figures are for 1999 except US (1997); UK, Austria (1998); Finland, Japan, Spain, Portugal (2000).

◆ www.cordis.lu/rtd2002/indicators

◆ europa.eu.int/comm/research/era/sti_en.html